

A Study of the Molecular Pathways of Wood Formation using Peptide Mass Spectrometry

Robert Nilsson

Faculty of Forest Sciences

Department of Forest Genetics and Plant Physiology

Umeå

Doctoral Thesis

Swedish University of Agricultural Sciences

Umeå 2010

Acta Universitatis agriculturae Sueciae
2010:64

Cover: From transcription to cell wall synthesis
(photo: R. Nilsson & K. Olofsson)

ISSN 1652-6880
ISBN 978-91-576-7477-7
© 2010 Robert Nilsson, Umeå
Print: Arkitektkopia i Umeå

A Study of the Molecular Pathways of Wood Formation using Peptide Mass Spectrometry

Abstract

Wood consists essentially of secondary plant cell walls, composed of a cellulose and hemicellulose network impregnated with lignin. It is formed via the development of derivatives of a thin layer of cells (the vascular cambium) into xylem cells through cell division, expansion, secondary wall formation, lignification and, finally, programmed cell death. Throughout the formation of the resulting, complex matrix (wood) diverse proteins are involved. Hence, biosynthesis of wood is tightly regulated at the molecular level, primarily by transcriptional regulation, together with various post-transcriptional processes.

In the studies this thesis is based upon, a key technique for studying proteins, liquid chromatography coupled to mass spectrometry, was used to elucidate pathways of wood formation. First, proteins involved in the central transcriptional unit Mediator, regulating most essential processes in plants, were identified in arabidopsis. Mediator subunits that most likely provide secondary cell wall activity are differentiated paralogous subunits located in the tail part of the complex. Point mutations in the Mediator tail subunit, Med5#2 affect secondary cell wall lignification. Moreover, control of xylem expansion may be indirectly regulated by Med25. A strategy to integrate transcript-, protein- and metabolite-data was then developed, using data acquired from analyses of poplar mutants with perturbed wood development. The results showed that lignin biosynthesis is heavily affected at both transcript and protein levels in the mutant with the strongest phenotypic deviations. Interestingly, when transcript levels are decreased, the protein levels are increased. Finally, the front of cell wall synthesis at poplar plasma membranes was dissected. A high coverage of proteins known to be involved in the cell wall synthesizing machinery and associated components was captured, as well as several potential new ones.

Keywords: proteomics, transcription, xylem, mediator, regulation, plasma membrane, development, arabidopsis, poplar, *Populus*, mass spectrometry

Author's address: Robert Nilsson, SLU, Department of Forest Genetics and Plant Physiology, 901 83 Umeå, Sweden
E-mail: Robert.Nilsson@genfys.slu.se

Varför ska man göra lätt för sig, när det är så lätt att göra svårt för sig.
Ronny Eriksson

Contents

List of Publications	7
Abbreviations	9
1 Introduction	11
1.1 Wood formation	12
1.1.1 Cell differentiation and physiology	12
1.1.2 Biochemical composition of wood	14
1.2 The central dogma and beyond	17
1.2.1 Transcriptional regulation	17
1.2.2 Alternative splicing of primary transcripts	20
1.2.3 Post translational regulation	20
1.3 Genes in the pathway of wood formation	21
1.3.1 Transcription factors and hormones	21
1.3.2 Wall building enzymes and supporting scaffolds	24
1.4 Functional genomics	29
2 Objectives	31
3 Material and methods	33
3.1 Liquid chromatography-mass spectrometry	33
3.2 Plant material and Protein extraction	33
3.2.1 Isolation of protein complex from cell culture	36
3.2.2 Isolation of water-soluble proteins from xylem tissue	37
3.2.3 Isolation of plasma membrane proteins	38
3.3 Protein digestion	39
3.4 Nanoflow liquid chromatography	39
3.5 Electrospray ionization	40
3.6 Time-of-flight mass spectrometry	41
3.7 Peptide fragmentation	42
3.8 Bioinformatics	44
3.8.1 Preprocessing of raw data	44
3.8.2 Databases and search engines	44
3.8.3 Bidirectional multivariate regression	47
3.8.4 Visualizing bioinformatics data	48

4	Results and discussion	51
4.1	Composition of the plant transcriptional regulator	51
4.2	Correlation between transcripts, proteins and metabolites	56
4.3	Components of wood formation in the plasma membrane	59
5	Conclusions and future perspectives	65
6	Sammanfattning	69
	References	71
	Acknowledgements	81

List of Publications

This thesis is based on the work contained in the following papers, referred to by Roman numerals in the text:

- I Backstrom, S., Elfving, N., **Nilsson, R.**, Wingsle, G., and Bjorklund, S. (2007) Purification of a plant mediator from *Arabidopsis thaliana* identifies PFT1 as the Med25 subunit. *Molecular Cell* 26, 717-729.
- II Bylesjo, M[★], **Nilsson, R[★]**, Srivastava, V., Gronlund, A., Johansson, A. I., Jansson, S., Karlsson, J., Moritz, T., Wingsle, G., and Trygg, J. (2009). Integrated Analysis of Transcript, Protein and Metabolite Data To Study Lignin Biosynthesis in Hybrid Aspen. *Journal of Proteome Research* 8, 199-210.
- III **Nilsson, R.**, Bernfur, K., Gustavsson, N., Bygdell, J., Wingsle, G., and Larsson, C. (2010). Proteomics of Plasma Membranes from Poplar Trees Reveals Tissue Distribution of Transporters, Receptors, and Proteins in Cell Wall Formation. *Molecular & Cellular Proteomics* 9, 368-387.

★ To be considered joint first authors

Papers are reproduced with the permission of the publishers.

The contribution of Robert Nilsson to the papers included in this thesis was as follows:

- I Planned identification analysis jointly with co-authors, extracted and analysed peptides using mass spectrometry, evaluated the data, and assisted in the writing of the paper.
- II Planned design of experiments, extracted proteins and metabolites, analysed peptides using mass spectrometry, evaluated the data, and wrote the paper jointly with co-authors.
- III Planned design of protein identification experiments, extracted and analysed peptides using mass spectrometry, evaluated the data, and wrote the paper jointly with co-authors.

Abbreviations

4CL	4-coumarate:CoA ligase
AMBIC	ammonium bicarbonate
Ara	arabinose
At	<i>Arabidopsis thaliana</i>
C3H	4-coumarate 3-hydroxylase
C4H	cinnamate-4-hydroxylase
CAD	cinnamyl alcohol dehydrogenase
CAlc5H	coniferyl aldehyde 5-hydroxylase
CAZyme	carbohydrate-active enzyme
CCoAOMT	caffeoyl-CoA O-methyltransferase
CCR	cinnamoyl-CoA reductase
CE	carbohydrate esterase
CESA	cellulose synthase
CID	collision-induced dissociation
COMT	caffeic acid/5-hydroxyconiferaldehyde O-methyltransferase
CSC	cellulose synthase complex
DNA	deoxyribonucleic acid
DoE	Design of Experiments
EMS	ethyl methanesulfonate
EST	expressed sequence tag
Fuc	fucose
Gal	galactose
GalUA	galacturonic acid
GC	gas chromatography
GH	glycoside hydrolase
Glc	glucose
GlcUA	glucuronic acid
GME	GDP-D-mannose 3,5-epimerase

GPI	glycosylphosphatidyl inositol
GT	glycosyl transferase
GTF	general transcription factor
HCT	<i>p</i> -hydroxycinnamoyl-CoA:quinic shikimate <i>p</i> -hydroxycinnamoyltransferase
HTM	heat and trapping module
IEF	isoelectric focusing
LC	liquid chromatography
Man	mannose
miRNA	microRNA
mRNA	messenger RNA
MS	mass spectrometry
MS/MS	tandem mass spectrometry
NAD	nicotinamide adenine dinucleotide
O2PLS	bidirectional OPLS
OPLS	orthogonal projections to latent structures
PAL	phenylalanine ammonia-lyase
PCA	principal component analysis
PCR	polymerase chain reaction
PIC	preinitiation complex
pre-mRNA	precursor mRNA
Ptr	<i>Populus trichocarpa</i>
Ptt	<i>Populus tremula x tremuloides</i>
Q-ToF	quadrupole time-of-flight
RG	rhamnogalacturonan
Rha	rhamnose
RHM	UDP-L-rhamnose synthase
RNA	ribonucleic acid
RNAi	RNA interference
RNAP	RNA polymerase
rRNA	ribosomal RNA
SAD	sinapyl alcohol dehydrogenase
SDS-PAGE	sodium dodecyl sulfate polyacrylamide gel electrophoresis
siRNA	short interfering RNA
UGE	UDP-D-glucose 4-epimerase
UPLC	ultra performance liquid chromatography
UXS	UDP-D-apiose/xylose synthase
WT	wild-type
Xyl	xylose

1 Introduction

55% of Sweden's surface area is covered by productive forest (Anon., 2008). Although the growth of this dense forest is much less rapid than that of its tropical counterparts, its wood can be profitably refined by using advanced management, harvesting, and transport techniques. When refining wood, the quality and value of the final products are determined by the composition of the major cell-wall polymer (a blend of cellulose, hemicellulose and lignin). Differences in the precise constitution of the hemicellulose mixture aside, the normal composition of hardwoods (such as poplar, birch, or oak) is comparable to that of softwoods such as pine and spruce (Pauly & Keegstra, 2008). Traditional breeding and genetic engineering can be used to alter the constitution and composition of the wood and thereby facilitate its refinement. However, such alterations may compromise the biological functionality of the cell wall by reducing its pathogen resistance, its ability to withstand mechanical stress, or its capacity for intracellular transport or cell signaling; such potential problems must be addressed. Field trials (or mixed field/greenhouse studies) are necessary to characterize the new phenotypes and expose pleiotropic effects. Potentially undesirable side-effects can be minimized in various ways. First, even relatively minor changes in wood chemistry can result in dramatic improvements to the refining process; one may not need to significantly disrupt the existing functionality of the system to achieve desirable results. Second, benefits can be sometimes obtained even if the altered trees constitute only a small fraction of the total feedstock, and they can be propagated in greenhouses or restricted fields. An alternative method to avoid disrupting existing traits that are important in survival and fitness would be to allow the tree to grow normally and only induce the desired changes shortly before it is to be harvested. An inducible gene expression system could be used to turn on wood modifying processes at the desired

time in a specific cell type or tissue. At present, there are only a few such systems that are both suitable for field trials and commercially viable (Corrado & Karali, 2009). Basic knowledge of how wood formation is regulated can also be applied in the development of improved methods for the pretreatment of feedstocks for the pulp and paper industries, for biofuel production and for other renewable materials. Currently, most research on trees is performed on the angiosperm poplar (*Populus spp.*) because of its relatively small genome compared to other trees. In addition, poplar is an attractive model system because its genome has been sequenced, various tools for studying it have been developed, and its growth and vegetative propagation are both rapid (Jansson & Douglas, 2007; Taylor, 2002). Despite its small genome, poplar has a more complex wood structure than do softwoods, in that it has additional vessel elements. The relatively close phylogenetic relationship between poplar and the primary plant model system, arabidopsis (*Arabidopsis thaliana*), makes comparative studies possible. Paper I describes a study of a gene regulation system that is common to both poplar and arabidopsis, and to plants in general, using arabidopsis as the model system. Many of arabidopsis' genes have been well characterized; because of its rapid generation time, it has also lost some redundant genes that are essential in other plants (Jansson & Douglas, 2007). The remainder of the work described in this thesis (papers II and III) focused on studying wood formation, using poplar as the model system. In this context, it should be noted that by using a combination of short- and long-day conditions, in conjunction with regular removal of the inflorescence stems, arabidopsis can be forced to produce extensive secondary xylem (Demura & Fukuda, 2007).

1.1 Wood formation

The formation of wood (xylogenesis) involves cell division, cell expansion, cell wall thickening and programmed cell death (PCD). Additionally, the deposition of an aromatic polymer (lignification) in the secondary cell wall makes it rigid and impermeable. After this extensive synthesis and deposition, the mature wall can represent over 95% of the cell's dry weight, with the cell wall proteins accounting for only 5-10% of this (Brett & Waldron, 1996; Cassab & Varner, 1988).

1.1.1 Cell differentiation and physiology

Wood formation in the trunk starts at the shoot apical meristem (SAM) of the young plant. The shoot apical meristems develop a cylindrical lateral meristem called the vascular cambium that is able to generate all of the cell

types necessary for wood formation. Wood formation occurs simultaneously and in parallel in the roots and the branches. The vascular cambium contains stem cells known as the initials, as well as partially differentiated phloem and xylem mother cells. There are two types of initials. Radial initials give rise to ray cells, which are essential for the translocation of nutrients between phloem and xylem and the storage of nutrients. Fusiform initials, by contrast, form xylem and phloem cells. The xylem cells are further subdivided into tracheary elements, vessel elements (only found in angiosperms), xylem fibres and axial parenchyma.

The earliest stages of wall formation (cell plate formation) can be monitored during cell division. The cell plate grows from the centre of the cell out towards the pre-existing cell walls, and fuses with them. The zone in which plate formation occurs is called the phragmoplast; it possesses an extensive network of microtubules that are involved in the transport of vesicles from the endomembrane system. The vesicles carry the proteins and other materials that are used to build the cell plate, and lipids to generate its plasma membrane. At the same time, channels (plasmodesmata) form around some of the microtubules linking the two nascent daughter cells. These microtubules may also form a desmotubule, a tube of membrane that spans the plasmodesma and is continuous with the endoplasmic reticulum. As cytokinesis proceeds, the phragmoplast loses definition and the cell plate joins with the parental cell wall to form the middle lamellae. The separated daughter cells deposit further material to generate a primary wall. The structurally-important cellulose microfibrils are usually randomly or longitudinally oriented in the primary wall.

The cells expand during the primary wall stage. Fibers and axial parenchyma cells expand primarily in radial direction while vessel elements also exhibit substantial tangential growth. Fiber elongation is achieved by intrusive tip growth that requires wall biogenesis and dissolution of the middle lamellae between neighboring cells. Cell wall enlargement is regulated by the cell's turgor pressure, growth substances that alter the pH of the wall, and by the turnover and cross-linking of wall components, mediated by enzymes and various ligands.

As the xylem cells reach their final size, secondary wall formation takes place. Three distinct layers, denoted S1, S2 and S3, form inside the primary wall. The S1 layer is formed first. Initially, its microfibril angle is approximately perpendicular to the cell's long axis; as the layer grows, this angle increases in a clockwise direction (as seen from the cell lumen), eventually reaching the longitudinal arrangement that characterizes the thickest S2 layer. In poplar the S2 layer is proportionally thinner in the

vessel elements than in the fibres. Finally, the microfibril angle undergoes a second clockwise increase, restoring the transverse arrangement in the S3 layer; overall, the fibrils describe a transverse helix spanning the three layers. The changes in microfibril orientation during the formation of successive cell wall layers are always paralleled by the reorientation of cortical microtubules. An additional tertiary layer called the protective layer is formed within the contact ray cells after autolysis of the neighboring vessel element (Déjardin *et al.*, 2010; Mellerowicz *et al.*, 2001; Brett & Waldron, 1996; Larson, 1994).

Distinctive structures are formed in reaction wood, which is generally formed in xylem tissue in response to mechanical stress of the stem. In poplar, the reaction wood takes the form of tension wood; its formation is characterized by an increased rate of cell division on the tension wood side, and reduced wood formation on the opposite side of the stem. On the tension wood side, the vessels are reduced in size and density and the fibers form a gelatinous layer (G layer) over the inner side of the secondary cell wall (Mellerowicz *et al.*, 2001).

The last stage of wall formation is programmed cell death. At this stage the tonoplast (vacuole membrane) ruptures and releases hydrolytic enzymes into the cytoplasm to degrade the cell's contents. The cell walls are left intact, and constitute the wood.

The process of cell wall lignification occurs in parallel with the final stages of xylem cell differentiation but is not directly connected to programmed cell death. Lignified parenchyma cells remain alive for several years (Déjardin *et al.*, 2010). The extent of lignification can vary significantly among cells within a tissue; the most extensive deposition is normally observed in vessels (Donaldson *et al.*, 2001). Lignification proceeds in different phases. Initially, following the initiation of S1, the corners of the cell walls become lignified, along with the middle lamella. After the S2 layer is completed, lignification proceeds further, and massive deposition occurs when the S3 layer is finished. Ultimately, the concentration of lignin within the wall is highest in the middle lamella and cell corners, and is relatively low in the S2 layer. However, the overall lignin content of the secondary wall is greater than that of other regions due to its larger volume (Boerjan *et al.*, 2003).

1.1.2 Biochemical composition of wood

Wood is a composite material with two phases, a microfibrillar phase and a matrix phase. The microfibrillar phase has a high degree of crystallinity and a homogenous chemical composition: it consists of cellulose microfibrils. The

chemical composition of the non-crystalline matrix phase is more complex. It consists of a variety of polysaccharides (hemicelluloses), proteins and phenolic compounds (lignin)(Brett & Waldron, 1996).

The most abundant component of the secondary wall of poplar is cellulose (43–48%), while pectin is the most abundant component of the primary wall (~47%). The cell plate is probably made of callose when it is first formed, but is re-made from cellulose at later developmental stages (as observed in tobacco) (Mellerowicz *et al.*, 2001). The biopolymer compositions of the primary and secondary cell walls of poplar are shown in Figure 1.

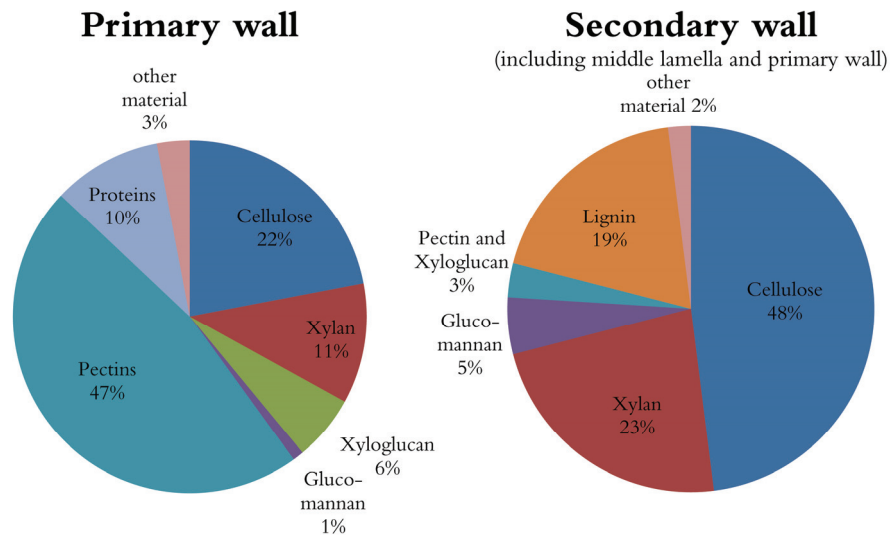


Figure 1. Biopolymer composition (% dry weight) of the primary and secondary cell walls of poplar. Adapted from Mellerowicz *et al.* (2001).

The detailed compositions of the major biopolymers (in terms of their constituent monomers) in the primary and secondary cell walls of poplar are compiled in Table 1.

Table 1. *Monomer composition of the main components of the primary and secondary cell walls of poplar, from Takahashi Schmidt thesis (Schmidt, 2008).*

Polysaccharides or lignins	Occurrence	Major monomers	Backbone	Substituents
Cellulose	Primary and secondary cell wall	β -Glc	β -1,4-Glc	Unbranched.
Xyloglucan	Mainly primary cell wall	β -Glc, α -Xyl, β -Gal, α -Fuc	β -1,4-Glc	α -Xyl-1,6- β -Glc, β -Gal-1,2- α -Xyl-1,6- β -Glc, α -Fuc-1,2- β -Gal-1,2- α -Xyl-1,6- β -Glc. β -Gal is partially acetylated.
Glucuronoxylan	Mainly secondary cell wall	β -Xyl, α -GlcUA	β -1,4-Xyl	α -GlcUA-1,2- β -Xyl. Some α -GlcUA are as 4-O-methyl ether. Some β -Xyl are acetylated.
Glucomannan	Mainly secondary cell wall	β -Glc, β -Man	β -1,4-Glc/ β -1,4-Man (ratio 1:2)	Unbranched.
Pectin Homogalacturonan	Primary cell wall	α -GalUA	α -1,4-GalUA	Unbranched. Some α -GalUA are as methyl ester.
Pectin RG-I	Primary cell wall	α -GalUA, α -Rha, β -Gal, α -Ara, α -Fuc	α -1,4-GalUA- α -1,2-Rha	Rich in α -Ara and/or β -Gal, attached to O-4 of α -Rha.
Pectin RG-II	Primary cell wall	Eleven different glycosyl residues. Usual sugars are α -GalUA, β -Rha, α -Gal, α -Fuc, α -Rha, β -GalUA, α -Ara, β -GlcUA.	At least eight α -1,4-GalUA	Two structurally distinct disaccharides attached to C3 and two oligosaccharides attached to C2 of the backbone.
Lignins	Middle lamella, primary cell wall (higher conc.) and secondary cell wall (lower conc.) during secondary cell wall stage.	Guaiacyl (G) and syringyl (S) units and traces of <i>p</i> -hydroxyphenyl (H) phenylpropanoid units.		

1.2 The central dogma and beyond

The central dogma of molecular biology states that sequential information in DNA is transferred by transcription to RNA and then by translation to protein in a unidirectional process (Crick, 1970). These processes have been studied extensively since the dogma was first articulated by Francis Crick in 1958. Recent work has shown that these processes are much more complex than the dogma implies, and that many more actors are involved than was initially thought, including microRNAs (miRNAs) and short interfering RNAs (siRNAs) (Carthew & Sontheimer, 2009; Vaucheret, 2006). In addition, DNA carries numerous regulatory elements, both within and outside genes and promoter regions; these regulatory elements are the focus of much current research. The proteins involved in these processes are often incorporated into large and dynamic multi-protein complexes (Sikorski & Buratowski, 2009). The most important enzyme complexes in the transcription of genomic DNA to RNA are the RNA polymerases (RNAPs).

1.2.1 Transcriptional regulation

In eukaryotes, there are three nuclear polymerases: RNAP I, which transcribes the ribosomal RNA (rRNA); RNAP II, which handles the bulk of the transcription of genes to precursor messenger RNA (pre-mRNA) and also that of most small nuclear RNA (snRNA) and miRNA; and RNAP III, which transcribes genes that encode short structural RNAs, including transport RNAs (tRNAs) and 5S rRNA. Plants have two additional polymerases, RNAP IV and V, which mediate siRNA-directed DNA methylation and gene silencing (Ream *et al.*, 2009; Wierzbicki *et al.*, 2009; Onodera *et al.*, 2005).

RNAP II transcribes all protein-coding genes. It consists of 12 proteins in humans, yeast and plants (Ream *et al.*, 2009; Myer & Young, 1998). To initiate transcription, RNAP II needs to bind to a number of other proteins to form the preinitiation complex (PIC). Minimally, transcription *in vitro* requires the presence of the basal initiation factors TFIIA, TFIIB, TFIID, TFIIIE, TFIIF, and TFIIF. These components are called general transcription factors (GTFs). The Mediator protein complex transmits information from regulatory DNA elements to core promoters and the basal machinery (Kim *et al.*, 1994; Koleske & Young, 1994) (Figure 2). There is some debate over whether Mediator should also be classified as a GTF (Sikorski & Buratowski, 2009).

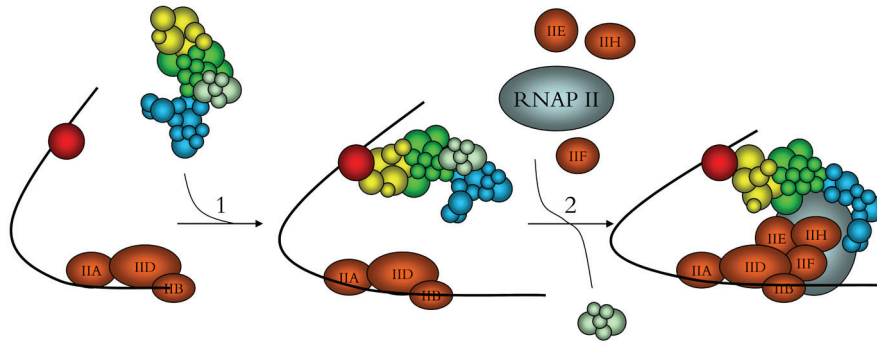


Figure 2. A possible mechanism for Mediator-activated transcription. In this scheme, a transcription factor (red) binds to its target site and recruits the intact form of Mediator (1). [It is not known whether all of the subunits of Mediator are recruited at all promoters]. The Mediator could then facilitate the recruitment of RNAP II and gather the PIC (2). This could be accompanied by dissociation of the MED13–MED12–CDK8–CycC module (grey) and other modifications (e.g. MED26 binding), perhaps in conjunction with the initiation of transcription. Adapted from Malik and Roeder (2005).

Mediator was first identified as being necessary for transcriptional activation *in vitro* in yeast (*S. cerevisiae*) (Flanagan *et al.*, 1991; Kelleher *et al.*, 1990). The yeast Mediator complex consists of 25 proteins, comprising four subdomains: head, middle, tail and kinase (Asturias *et al.*, 1999). The kinase domain is only associated with free Mediator (i.e. Mediator that is not bound to RNAP II) (Samuelsen *et al.*, 2003). The head and middle domains bind to the C-terminal domain of RNAP II while the tail interacts primarily with DNA-bound transcription factors. When so bound, the Mediator can act as a molecular switchboard and mediate signals from hormones, receptors and transcription factors (van den Berg *et al.*, 2010; Khidekel & Hsieh-Wilson, 2004). A key feature of Mediator is its ability to regulate both positive and negative signalling pathways. Different subunits have been shown to act antagonistically on the same genes in yeast. Several Mediator subunits in yeast are essential, primarily those located in the head domain (van de Peppel *et al.*, 2005). Metazoan Mediators have been identified in *D. melanogaster*, *C. elegans*, mouse, and human cells. They share 22 subunits with the yeast Mediator (Myers & Kornberg, 2000) but also contain eight subunits (Med 23–Med 30) that are not found in *S. cerevisiae*. Finally, Mediator has also been purified from the fission yeast *S. pombe*, which contains a somewhat smaller Mediator. However, 19 of the shared metazoan and *S. cerevisiae* Mediator subunits and one of the metazoan-specific Mediator subunits are conserved in the *S. pombe* Mediator. These results led to a unified nomenclature for Mediator subunits in all species (Bourbon *et*

al., 2004). Before our publication (paper I), there were no reports of the biochemical identification of Mediator in plants.

Analysis of the genomic sequence of the model plant arabidopsis revealed that only 8%–23% of the proteins involved in transcription are related to genes found in other eukaryotic genomes, indicating that many plant transcription proteins evolved independently. In contrast, 48%–60% of the arabidopsis genes that encode translation proteins have counterparts in other eukaryotic genomes, indicating that the functions of the corresponding proteins are highly conserved. It is notable that >5% of the arabidopsis genome codes for ~1 700 transcription factors (TFs), about 41% of which are from families specific to plants. It is therefore likely that coregulators that function at the interface between the rapidly evolving regulatory TFs (activators and repressors) and the more conserved general transcription machinery should contain both conserved and species-specific domains. Arabidopsis also contains several tandem gene duplications and large scale duplications on different chromosomes, accounting for >60% of the genome (Blanc *et al.*, 2000; Riechmann & Ratcliffe, 2000; Lin *et al.*, 1999; Mayer *et al.*, 1999). Some duplicated sequences subsequently underwent rearrangement and divergent evolution, but still, up to 40% of the arabidopsis genome seems to consist of pairs of closely related sequences.

Regulatory TFs are nuclear proteins that can regulate transcription in particular tissues and in response to cellular signals. The general view of regulatory TFs is that they bind to short DNA sequences (consensus motifs) at specific sites, typically in the promoter region of the relevant gene, and also to enhancers, which may be situated upstream or downstream from the transcriptional unit, or even within the unit itself. The transcription-affecting DNA motifs at the binding site are called regulatory elements (Latchman, 1998); it should be noted that the absence of a consensus motif does not preclude TF binding. Various mechanisms for TF activation in the absence of a proximal consensus site have been suggested: binding at a distal site that contains a consensus motif and then activating the site in question through protein–protein interactions; ‘piggyback’ binding that is mediated by protein–protein interactions with a second factor and that does not involve the DNA-binding domain of the first factor; and assisted binding to a site that is similar to the consensus site, which is enhanced by protein–protein interactions with another site specific DNA-binding factor or with a specifically modified histone. Factors that can be recruited to the genome in multiple ways are able to participate in multiple signaling pathways. The binding of a single regulatory TF can be essential, but not sufficient, for regulation of a specific gene. Overall, gene regulation in eukaryotes seems to

be heavily dependent on the cooperative and communal action of regulatory TFs (Farnham, 2009).

1.2.2 Alternative splicing of primary transcripts

Primary transcripts (pre-mRNA) generated by RNAP II can be spliced in various ways and therefore generate multiple templates (mRNA) from a single gene. In the nucleus the pre-mRNA are capped, spliced and polyadenylated before being exported to the cytosol for subsequent translation (Reddy, 2007). In plants, 20% of the genes undergo alternative splicing; the functionality of most of these has not yet been characterized (Barbazuk *et al.*, 2008). The role of alternative splicing as a developmental and/or tissue-specific regulator is a topic of current interest (Srivastava *et al.*, 2009).

1.2.3 Post translational regulation

Transcripts exiting the nucleus at the nuclear pores are translated, relocalized or stored before ultimately being degraded (Bailey-Serres *et al.*, 2009). Post-translational modifications (PTMs) of proteins are of great importance in transcription and gene regulation. The covalent binding of various molecules to the amino acids of proteins makes the proteome two to three orders of magnitude more complex than would be expected solely on the basis of the number of genes in the genome and the various alternative splicings they can undergo (Walsh *et al.*, 2005). Around 300 different PTMs (Witze *et al.*, 2007) have been identified; they affect the target protein's localization, stability, and interactions, and thus, its activity. In general, only a subset of the translated proteins are modified; this complicates the search for regulatory proteins and complexes (Jensen, 2006). Furthermore, some PTMs are reversible, rapid, and highly dynamic. For instance, phosphorylation is a reversible PTM that is commonly involved in signal transduction. Kinases and phosphatases regulate the phosphorylation of their targets with temporal resolutions measured in minutes (Olsen *et al.*, 2006). Transcription of the genomic chromatin requires a cascade of PTMs of both the histones and the transcriptional machinery (Weake & Workman, 2008; Zhou *et al.*, 2008; Pfluger & Wagner, 2007).

1.3 Genes in the pathway of wood formation

Wood formation is subject to highly regulated genetic control, primarily at the transcriptional level. Transcription factors (TFs) are some of the first active genes in the pathways leading to wood formation. The following section provides an overview of the current models of the regulation of wood formation, from the cellular to the tissue level.

1.3.1 Transcription factors and hormones

The first set of TFs that control early vascular development are regulated by hormones and miRNA (Demura & Fukuda, 2007) (Figure 3). In this model vascular formation is initiated by the binding of the TF ARF5 to the DNA following the recruitment of the inhibitory protein IAA12 by the auxin-binding protein TIR-1. This causes the degradation of IAA12 via the ubiquitin proteasome pathway. Such regulation via proteolysis is rather slow, with a lag time of 10-15 min; other hormone (auxin) responses are typically detected within a few minutes (Demura & Fukuda, 2007). Downstream of ARF5, there are two classes of TFs that regulate tissue patterning and polarity. The first class comprises four genes encoding class III homeodomainleucine zipper (HD-ZIPIII) transcription factors: IFL1, ATHB8, ATHB9 and ATHB14. The second comprises three members of the GARP-type transcription factor family: KAN1, KAN2 and KAN3. The HD-ZIPIII and KAN families have a complex pattern of overlapping and antagonistic functions involving control of the stability of the HD-ZIPIII transcripts by miRNA. Furthermore, it is possible that brassinosteroids (BRs) may be involved in up-regulating the HD-ZIPIII genes to promote xylem formation.

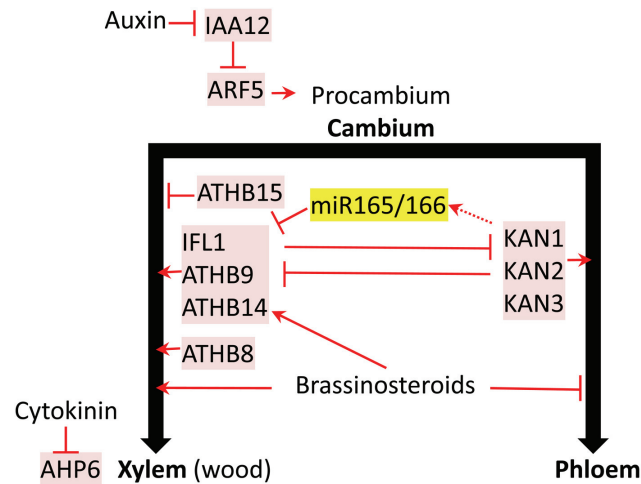


Figure 3. Control of the early stages of tissue formation by transcriptional regulators and hormones. Arrows indicate activation; T-bars indicate inhibition. Adapted from Demura and Fukuda (2007).

It was recently shown that a group of NAC domain TFs (PtrWNDs) may be key regulators of the cellulose, xylan and lignin pathways in poplar. The PtrWNDs are most likely orthologs of the arabidopsis AtSND1, a master transcriptional switch which activates a network of other TFs (Figure 4) involved in the control of the pathways that form the three main components of wood (Zhong *et al.*, 2009).

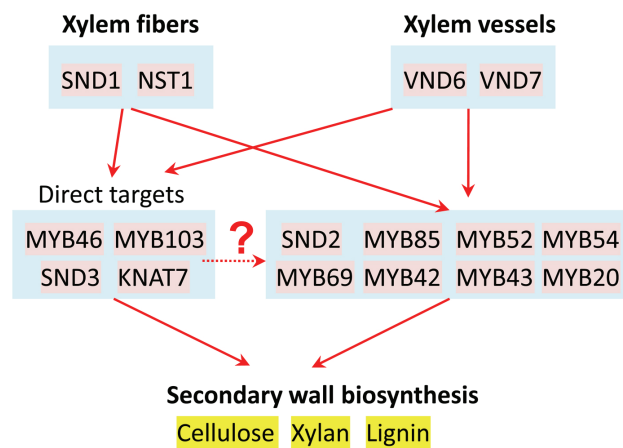


Figure 4. Transcriptional regulation of secondary wall formation in arabidopsis. Arrows indicate activation. Adapted from Zhong *et al.* (2008).

The TFs downstream of AtSND1 are required for normal secondary wall biosynthesis in arabidopsis. *In situ* hybridization localizes expression of all PtrWND genes to developing vessels and fibers. Additionally, PtrWND transcripts are observed in ray parenchyma cells, and a subset of the PtrWNDS are expressed strongly in phloem fibers. Overexpression of PtrWND2B and PtrWND6B (by the CaMV 35S promoter) in arabidopsis induces the downstream TFs associated with secondary wall formation as well as the biosynthetic genes for cellulose (AtCesA4, AtCesA7 and AtCesA8), xylan (AtIRX8, AtIRX9 and AtFRA8) and lignin (At4CL1 and AtCCaAOMT1). Moreover, ectopic deposition of cellulose, xylan and lignin is observed in PtrWND2B and PtrWND6B overexpressors. However, only about 20% of these transgenic plants exhibit the 'curly leaves' phenotype. In arabidopsis 35S-AtSND1, around 50% of the overexpressors have a characteristic visual phenotype in which the rosettes are small and the leaves are stunted and have severely upward-curling blades (Zhong *et al.*, 2006). It is thought that the 'curly leaves' phenotype is only observed in a few of the transgenic plants because of differences in the cells' competence to form secondary walls (Zhong *et al.*, 2009).

The R2R3-MYB family is a large group of TFs that are heavily involved in the transcriptional activation of secondary wall formation. There are 192 R2R3-MYB family members in poplar (*P. trichocarpa*). An analysis of transcript abundance in poplar indicated that 23 out of 180 R2R3-MYB encoding genes are involved in xylem differentiation. It has been suggested that R2R3-MYB proteins may function as regulators of processes that are limited to discrete cells, organs, or conditions (Wilkins *et al.*, 2009). Paper II describes a study investigating the role in wood formation of an R2R3-MYB TF which is upregulated in the latter stages of xylem formation (Karpinska *et al.*, 2004). According to the previously-suggested nomenclature (Wilkins *et al.*, 2009), this gene should be denoted PttMYB090 (in the article, it is referred to as PttMYB21a). Phylogenetic analyses suggest that this gene is one of two MYBs in poplar (PtrMYB090 and PtrMYB161), both of which are orthologs of AtMYB52 from arabidopsis. PttMYB090 is rapidly (0.5–3h) induced following bending (tension) of the stem, suggesting that it may be involved in the early stages of tension wood regulation. Furthermore, antisense constructs of PttMYB090 increases the expression of CCoAOMT (caffeoyl-CoA O-methyltransferase), an enzyme involved in lignin biosynthesis (Karpinska *et al.*, 2004). It was recently shown that expression of AtMyb52 activates all three of the major secondary wall biosynthetic pathways. However, this

alone was not sufficient to induce ectopic deposition of secondary walls (Zhong *et al.*, 2008).

Prior to the identification of Mediator in *Arabidopsis* as described in paper I, the mediator subunit AtMed33b (AtMed5 using the updated nomenclature (Bourbon, 2008); accession number At2g48110) was identified during a screen for altered phenylpropanoid compounds in EMS-mutagenized plants. By means of a combination of mapping and sequencing, abnormally low levels of phenylpropanoid species were linked to amino acid substitutions at conserved positions in At2g48110. Plants bearing the mutant allele, named *ref4*, have reduced levels of all of the major phenylpropanoids, including lignin, flavonoids, and sinapate esters. The unusual semidominant behaviour associated with the mutant phenotype is characteristic of a mutation in a TF, transporter, or a component of a signalling cascade. It has been shown that dominant mutant traits can arise from point mutations that increase or stabilize mRNA transcripts or the proteins they encode, preventing normal turnover. For example, the *atr1D* mutant exhibits upregulated transcription of tryptophan pathway genes due to stabilization of the ATR1 MYB transcription factor mRNA. Moreover, the *ref4* mutant could be rescued by a second point mutation exchanging a proline for a leucine located in a WW protein-interaction domain (DWPSPA). The WW domain is conserved in all At2g48110 plant homologs and is a submotif within a proline-directed serine kinase phosphorylation site (DWPSPA) (Stout *et al.*, 2008).

1.3.2 Wall building enzymes and supporting scaffolds

Carbohydrate-active enzymes (CAZymes) are responsible for the synthesis, modification, and degradation of the carbohydrate polymers that comprise the cell wall, both in terms of the bonds that comprise the polymers' backbones and those at branch points. These enzymes are also responsible for the glycosylation of proteins, lipids and nucleic acids. In poplar, CAZyme transcripts are most abundant in woody tissues (Geisler-Lee *et al.*, 2006). CAZymes are classified into different groups on the basis of their amino acid sequences. This is facilitated by their modular composition, featuring one or more catalytic domains and a variable number of independent substrate binding sites. Enzymes are assigned to a particular class on the basis of conserved active-site residues, molecular mechanisms, and protein folds (Davies & Henrissat, 2002). There are four classes of CAZymes; glycosyl transferases (GTs), glycoside hydrolases (GHs), polysaccharide lyases (PLs) and carbohydrate esterases (CEs). The CAZyme system provides a framework that is useful in the study of wood formation

in poplar, for which gene annotation is limited and is largely based on functional characterizations of analogous genes from arabidopsis.

Xylem-specific CAZymes in poplar have been identified using microarray analysis of narrow tissue sections taken from developing xylem. In total 25 GTs and 9 GHs were found to have secondary wall-specific expression patterns, but no PL or CE genes were significantly expressed (Aspeborg *et al.*, 2005).

The key cellulose-synthesizing enzyme complexes are located in the plasma membrane and contain at least three types of cellulose synthase (CesA) proteins (Somerville, 2006) belonging to the GT2 family. PttCesA4, PttCesA8-A, PttCesA8-B and PttCesA7-x (classification according to Kumar *et al.* (2009)) are particularly highly expressed during xylogenesis (Aspeborg *et al.*, 2005) and are therefore plausible candidate members of these complexes. The corresponding CesA homologs in arabidopsis are associated with secondary cell wall formation (Persson *et al.*, 2005). The transcript of the cellulose synthase-like gene, PttGT2A is very abundant at the onset of secondary cell wall formation, indicating that the protein it encodes may be a mannan or xylan synthase.

Several genes in the GT8 family of enzymes that catalyze the formation of α -glycosidic linkages are expressed at high levels during secondary wall formation. PttGT8A, PttGT8B, and PttGT8C may be self-glucosylating proteins that serve as primers for polysaccharide synthesis. PttGT8D, PttGT8E, PttGT8F, and PttGT8G may be specifically involved in pectin biosynthesis in the contact ray cells. Two members of the GT14 family, PttGT14A and PttGT14B are expressed in xylem and have plant-specific features; they await functional characterization. In animals, GT14 enzymes catalyze the transfer of β (1-6)-linked N-acetylglucosaminyl and β -linked xylosyl residues to proteins. PttGT31A and PttGT31B from the GT31 family are also expressed, but no plant GT31 has been functionally characterized (in animals, these enzymes form β -glycosidic linkages between α -linked nucleotide sugars). Two genes from the GT43 family (in animals, enzymes in this family are β -glucuronyltransferases), PttGT43A and PttGT43B, may be involved in fiber development. Four genes from the GT47 family are highly expressed; the first two, PttGT47A and PttGT47D, may be involved in the transfer of glucuronyl side chains to 4-O-methylglucuronoxylan in the secondary wall. Alternatively, the enzymes may be involved in the synthesis of rhamnogalacturonan II in ray cells. The second two, PttGT47B and PttGT47C, have not been characterized. One gene from the GT61 family, PttGT61A, is highly expressed during xylogenesis. The enzymes in subfamily A catalyze the transfer of Xyl from

UDP-Xyl to the core β -linked Man of N-linked oligosaccharides in glycoproteins.

The proteins encoded by genes from the GH9 family are typically endoglucanases which act on β (1-4)-glucan polymers (cellulose). The expression of PttCel9A is up-regulated during secondary cell wall formation. PttCel9A is probably a homolog of KOR1 in arabidopsis, which has been proposed to release cellulose chains from a putative sitosterol-glucoside primer or to be involved in the assembly or editing of cellulose microfibrils. PttXyn10 (a member of the GH10 family) is expressed in the xylem and may be involved in xylan hydrolysis or remodeling during xylogenesis. PttXET16L is a highly-expressed member of the GH16 family; it may have a role in the endohydrolytic degradation of xyloglucan. PttGH17 (a member of the GH17 family, which code for enzymes that degrade β (1-3)- or β (1-3)(1-4)-glucans) may have a role in callose degradation. PttGH19A from the GH19 family is expressed in a xylem-specific fashion, but its function is unknown. The GH28 family contains enzymes associated with pectin degradation and cell wall modification. Of these enzymes, PttGH28A is highly expressed and may be involved in the modification of the primary wall and the region between it and the developing secondary walls. The members of the GH35 family that have been characterized to date are β -galactosidases that act on various substrates, including arabinogalactans, galactolipids, and pectin. PttBGal35A and PttBGal35B are active during xylogenesis and may play a role in pit membrane adaptation and the hydrolysis of primary walls in the perforations, by degrading the galactan side chains of rhamnogalacturonan I or xyloglucan. Alternatively, their function may have to do with modifying arabinogalactan proteins. Members of the GH51 family have been shown to have α -L-arabinofuranosidase activity, releasing Ara from substrates such as arabinan, arabinoxylan, gum arabic, and arabinogalactan. PttGH51 is expressed during xylem formation and might act on homogalacturonan and the side chains of rhamnogalacturonan I.

Although the pathways involved in the biosynthesis of carbohydrate polymers in poplar have not yet been fully characterised, those involved in the biosynthesis of lignin are known in much greater detail. Lignin is a complex aromatic heteropolymer derived mainly from three monomers with differing degrees of methoxylation. These monolignols generate *p*-hydroxyphenyl (H), guaiacyl (G), and syringyl (S) phenylpropanoid units when incorporated into the lignin polymer. Angiosperm (poplar) lignin consists mainly of G and S units, with traces of H units. The monomers are synthesized from phenylalanine (derived from the shikimate biosynthesis

pathway) via phenylpropanoid and monolignol-specific pathways (Boerjan *et al.*, 2003). Recently, 95 phenylpropanoid gene models extracted from poplar (*P. trichocarpa*) were investigated. 18 genes that were considered likely to encode monolignol biosynthesis enzymes were found to be highly transcribed in xylem undergoing differentiation. The current model of monolignol biosynthesis involves 10 enzyme families (Shi *et al.*, 2010), as shown in Figure 5.

acid/5-hydroxyconiferaldehyde O-methyltransferase), CAD (cinnamyl alcohol dehydrogenase), SAD (sinapyl alcohol dehydrogenase).

Finally, several cytoskeleton genes are essential for the deposition and remodeling of the secondary cell wall. It is well established that microtubules have an important role in wall formation. The large family of tubulin genes in poplar contains some distinct isoforms that are expressed particularly strongly in tissues undergoing secondary cell wall thickening (Oakley *et al.*, 2007). The role of actin is less well understood. One model suggests that actin transports organelles containing cellulose synthase complex (CSC) around the cell, pausing at sites marked by transverse actin. The CSCs are incorporated into the plasma membrane at these sites, and are maintained beneath sites where wall synthesis is occurring by microtubules (Wightman & Turner, 2008).

1.4 Functional genomics

In the post-genomics era, functional genomics studies have largely focused on applying profiling techniques for the parallel monitoring of various aspects of cellular activity, such as transcription, protein synthesis, and the production of metabolites (Carrari *et al.*, 2006; Kolbe *et al.*, 2006; Rischer *et al.*, 2006; Hirai *et al.*, 2005; Tohge *et al.*, 2005; Clish *et al.*, 2004; Hirai *et al.*, 2004; Kleno *et al.*, 2004; Oresic *et al.*, 2004; Gygi *et al.*, 1999). This approach has become possible mainly due to the increasing availability of the instrumentation required for high-throughput characterization of biological samples, such as the microarray technology used in transcript profiling (Schena *et al.*, 1995) or the use of high performance chromatography in conjunction with mass spectrometry for peptide or metabolite profiling (de Hoffmann & Stroobant, 2001). The objective of such studies is to examine organisms as integrated systems of genetic, protein, metabolic, pathways and cellular events in order to achieve a higher level of understanding of the interplay between molecular and cellular components.

2 Objectives

The objective of the work described in this thesis was to understand the process of wood (cell wall) formation in terms of proteomics. Wood formation is a complex process involving a large set of genes. In an attempt to obtain a holistic view of the pathway, we have studied a range of processes, from the transcription of genes to the synthesis of the wall itself. Recent developments and improvements in nanoflow liquid chromatography and electrospray ionization mass spectrometry have made it possible to study proteins (via peptides) in unprecedented detail. It has become possible to validate hypothetical model pathways on the basis of expression analysis (transcripts) and to both identify individual gene products and quantify their levels of expression. Paper I describes the determination of the composition of the co-regulatory transcription complex called the “Mediator”. The Mediator is thought to act as a molecular switchboard for upstream transcriptional activators of various different pathways. The correlation between transcription and protein expression (including a subset of metabolites) is discussed in paper II. Finally, paper III describes the isolation and characterization of plasma membranes from various tissues of poplar, which contain the major components involved in cell wall synthesis. A detailed understanding of the molecular pathway of wood formation would make it possible to manipulate the composition of the wood and to fine tune the regulation of the synthetic flow for sustainable growth. This knowledge would be of use in academic research and in the industrial production of pulp, paper, biofuel, and other renewable materials from forestry.

3 Material and methods

3.1 Liquid chromatography-mass spectrometry

The first techniques for the analysis of proteins and peptides using mass spectrometry (MS) were only suitable for use with volatile samples that could be vaporized without undergoing thermal destruction. In 1988, John Fenn and co-workers made a major breakthrough when they invented the concept of electrospray ionization. The electrospray ion source made it possible to perform MS analysis on liquid samples, especially those which had been purified in advance by liquid chromatography (LC) (Whitelegge, 2008). In order to analyze complex protein extracts and their digests, different separation methods are required. It is mainly the huge dynamic range in protein abundance that makes sample processing necessary. To analyze proteins of low abundance, it is advisable to start by extracting only differentiated tissues, individual cell types, or specific subcellular compartments in which the target protein is relatively abundant. Alternatively, where applicable, affinity purification of individual proteins and their associated complexes often gives superior results. When purifying individual proteins, extracts are commonly separated on the basis of protein size by gel electrophoresis after complete denaturation, and on the basis of the proteins' isoelectric points using isoelectric focusing (IEF). Peptides from digested proteins are separated by liquid chromatography and analyzed using mass spectrometry.

3.2 Plant material and Protein extraction

Two plant model systems and several protein extraction methods were used in the work described in this thesis. First, arabidopsis was grown in cell

suspension cultures (paper I) and extracted in order to isolate a protein complex by means of ion exchange chromatography, immunoprecipitation and gel electrophoresis (Figure 7). Second, we studied hybrid aspen (*Populus tremula* × *Populus tremuloides*) clones grown in greenhouses (paper II) and in the open field (paper III). The studies described in paper II used wild-type (WT) T89 clones and two different transgenic lines bearing antisense constructs of PttMYB090 (PttMYB21a). The antisense plants were named G3 and G5, corresponding to the “21III” and “21V” lines described by Karpinska *et al.* (2004). The experiment was designed so as to enable the separate analysis of developmental and genotype effects (Figure 6); Design of Experiments (DoE) was also used to reduce the sample size (see the Bioinformatics section). Tissue containing differentiated xylem cells was ground in liquid nitrogen, and samples of this powder were extracted to isolate three specific types of biomolecules: proteins (from 20 mg of powder), total RNA (from 50 mg) and metabolites (from 10 mg). Highly water-soluble proteins were extracted using a variation on the method of Giavalisco *et al.* (2003), adapted for use with LC-MS rather than a gel system (Figure 8).

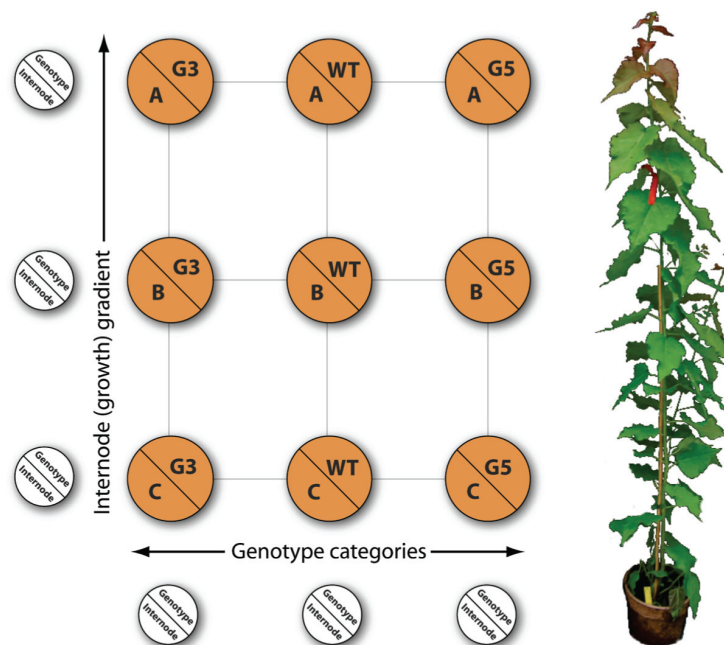


Figure 6. Overview of the study design. The different sample categories are shown, including three genotypes (G5, G3 and WT) and three internodes (A-C) on a 3 × 3 grid. The transcript, metabolite, and protein contents of samples from all 9 categories were measured. The smaller white circles explain the contents of the larger filled circles.

In the work described in paper III, the plasma membranes of fresh leaves and of cambium/phloem and xylem tissues were isolated by a combination of aqueous polymer two-phase partitioning and ultra centrifugation, after which their proteins were separated by gel electrophoresis (Figure 9).

3.2.1 Isolation of protein complex from cell culture

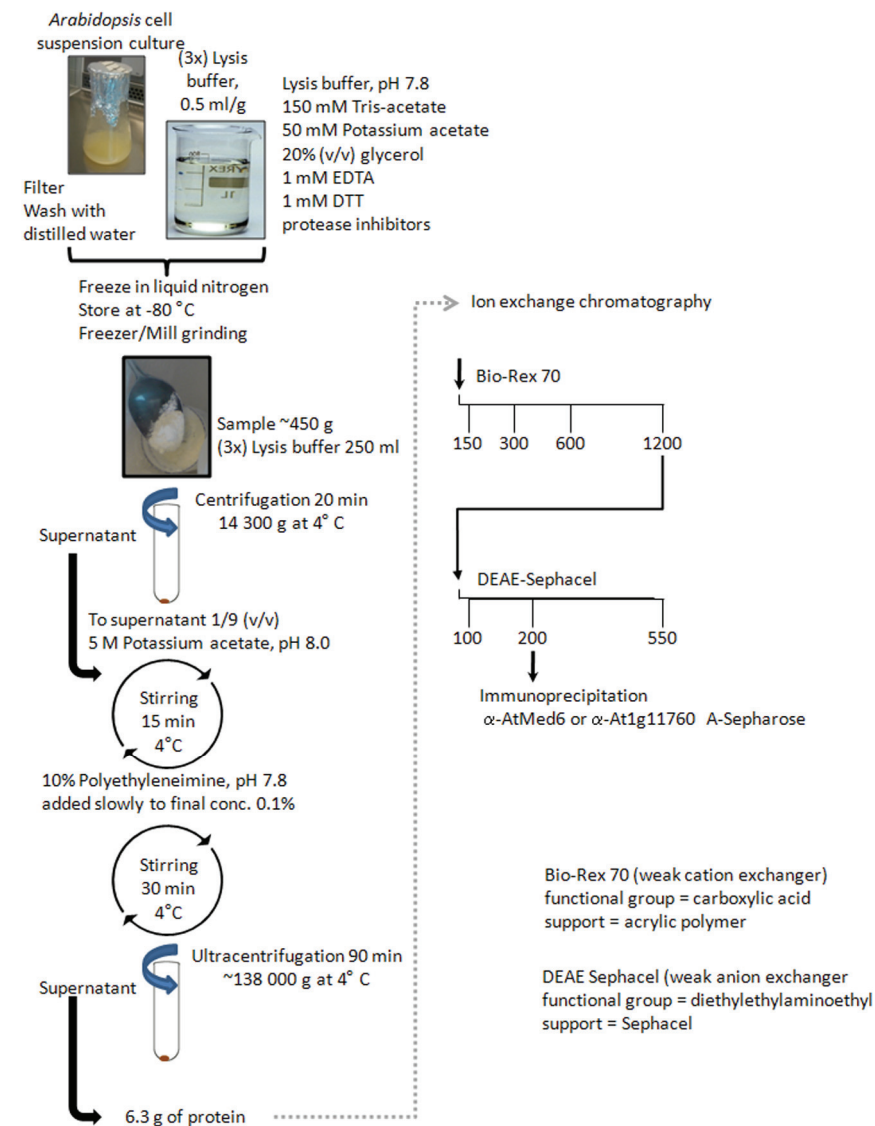


Figure 7. Overview of the procedure used when isolating protein complexes from cell cultures.

3.2.2 Isolation of water-soluble proteins from xylem tissue

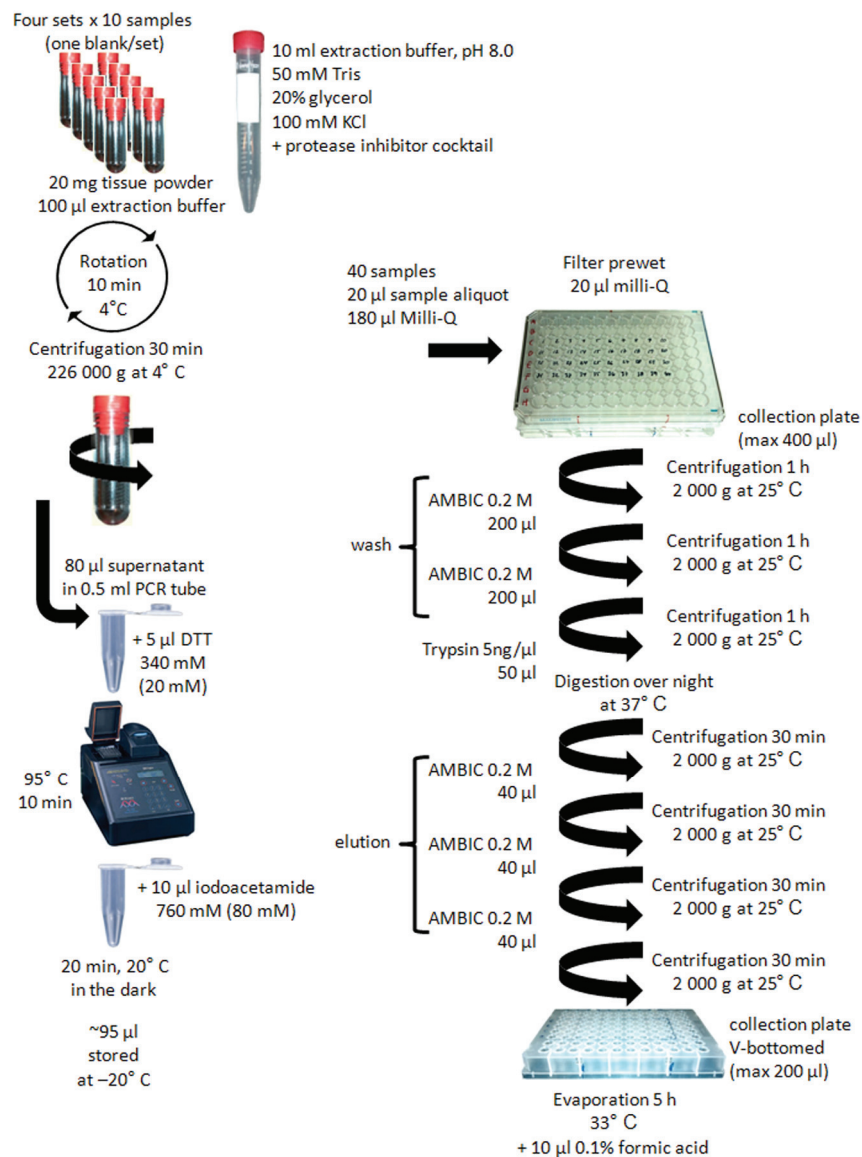


Figure 8. Overview of the procedure used when extracting water-soluble proteins from differentiated xylem tissue.

3.2.3 Isolation of plasma membrane proteins

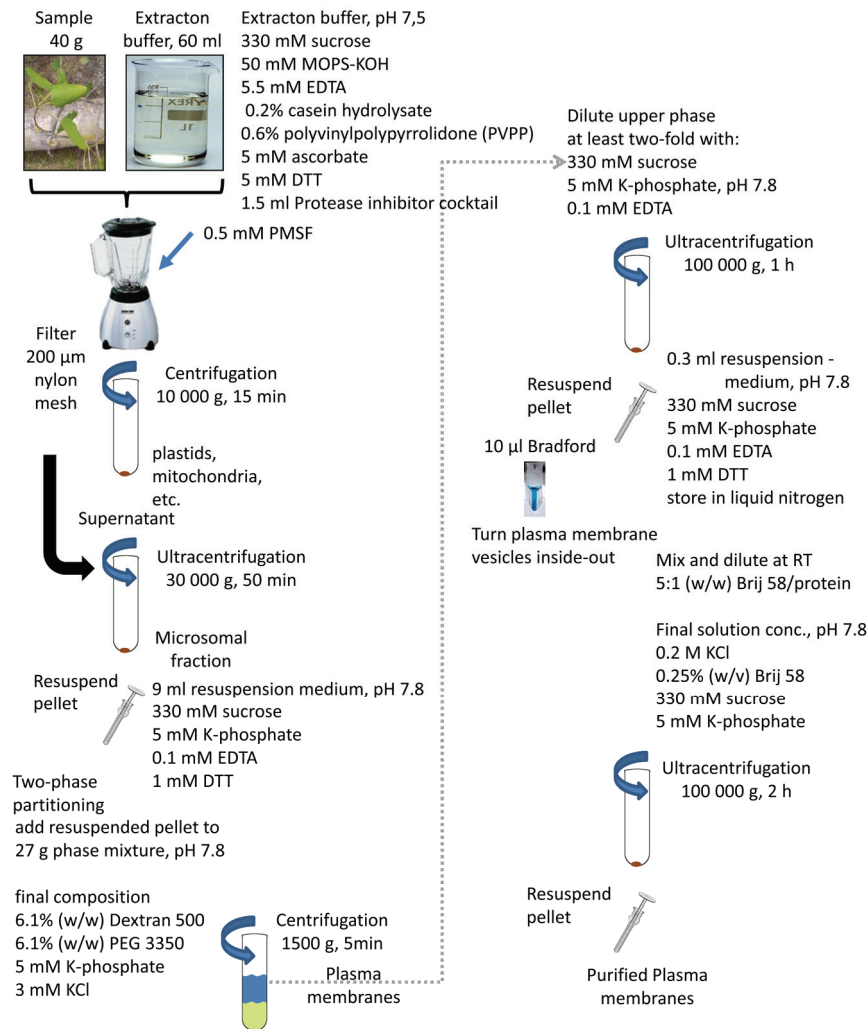


Figure 9. Overview of the procedure used when extracting plasma membranes from differentiated xylem tissue.

3.3 Protein digestion

Protein sequencing by MS/MS relies on the prior digestion of the proteins to peptides. We have performed digestion on proteins separated using gel systems and on some separated using non-gel based systems. In order to maximise the yields from gel systems, high gel surface areas are desirable, as this improves both rehydration and the recovery of peptides. However, there is a lower limit on the size of the gel pieces that can be used (and thus an upper limit on the possible gel surface area) because gel pieces smaller than 0.5 mm³ clog the pipet tips and cause losses of gel material (Havlis *et al.*, 2003). The duration of the incubation time with the enzyme is also critical: the penetration of trypsin into the gel matrix is controlled by diffusion, which is much slower than the swelling of the dehydrated gel. Consequently, proteases that are larger than trypsin (e.g., thermolysin) are much less efficient at in-gel digestion than in-solution digestion (Havlis *et al.*, 2003). The handling of proteins separated by SDS-polyacrylamide gel electrophoresis requires laborious techniques that can easily introduce contamination. Potential contaminants include dust (keratin) that adheres to plastic tubes by static electricity, detergents from sample buffers, and residual reagents used in sample treatment. Reduction/alkylation is a general method for sample treatment that usually improves separation and identification but can also reduce sensitivity (Borchers *et al.*, 2000). We omitted the reduction/alkylation step in the work described in paper III; the omission increased the sensitivity of the analysis and also considerably reduced the sample treatment time.

3.4 Nanoflow liquid chromatography

The complex peptide mixtures generated in these studies were separated using liquid chromatography (LC). The stationary phase of the column was functionalised with linear C₁₈ (n-octadecyl) chains, which give the column a high capacity and make it very hydrophobic. Peptides are retained on the column primarily due to the combined hydrophobicity of their individual amino acids (Krokhin *et al.*, 2004). Separation is carried out using a nano-scale flow of the liquid phase (water, acetonitrile and 0.1% formic acid) to achieve efficient electrospray ionization and thereby increase sensitivity. To facilitate the use of sample injection volumes in the 1-10 µl range, a pre-column was used to concentrate and wash the samples (at flow rates measured in µl/min). Thus, the sample is injected and washed on the

trapping column (Figure 10) with the HTM (heat and trapping module) valve open, allowing the excess eluent to flow into the waste container. The HTM valve is then closed, redirecting the flow to the analytical column to separate the sample at a flow rate of 200–600 nl/min.

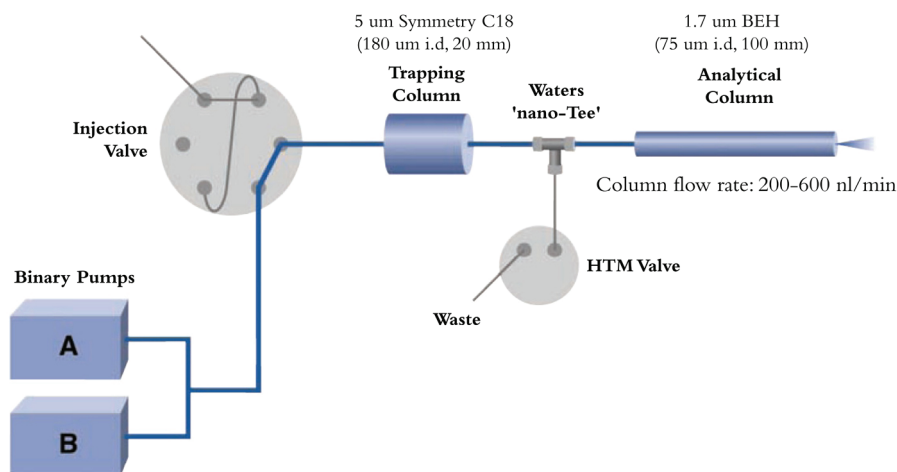


Figure 10. Setup of the nanoscale UPLC system. The injection valve allows for loading of the sample into the sample loop (10 µl) and, in conjunction with the HTM (heat and trapping module) valve, delivers it to the analytical column. The binary pumps provide a steady flow of the liquid phase (water, acetonitrile and 0.1% formic acid), mixed according to a user-specified gradient. The analytical column is housed in a heated sleeve for thermal control and enhanced reproducibility.

3.5 Electrospray ionization

In the electrospray ionization (ESI) of peptides, an acidic, aqueous peptide solution is sprayed through a small-diameter needle. A high positive voltage is applied to this needle to produce a Taylor cone from which droplets of the solution are sputtered. The protons from the acid give the droplets a positive charge, causing them to move away from the needle towards the negatively charged instrument. During the course of this movement, evaporation reduces the size of the droplets until the number and proximity of the positive charges split the droplet into a population of smaller, charged droplets. The evaporation process is often accelerated using a stream of heated nitrogen gas. The droplet-splitting is repeated until the small size and high charge of the droplets causes desorption of the protonated peptides into the gas-phase (Kinter & Sherman, 2000). The acidic conditions used in

electrospray ionization result in protonation of the peptides' basic functional groups, such as the N-terminal amine and the basic side chains of lysine (K), arginine (R) and histidine (H). As a result, multiply-charged peptide ions dominate the mass spectrum.

3.6 Time-of-flight mass spectrometry

A hybrid quadrupole time-of-flight (Q-ToF) mass spectrometer was used in all of the MS experiments discussed in this thesis. This instrument can also be classified as a tandem-in-space instrument, in contrast to instruments with only one mass analyzer, which perform tandem-in-time analysis. The operational principals of time-of-flight mass analysers are elegantly simple. Ions are accelerated by a fixed electric field to a kinetic energy (and thus a velocity) that is inversely proportional to their m/z (equation), and then travel at this velocity through a field-free area.

$$v = ((2 \star V \star z) / m)^{1/2}$$

m = mass in kg, v = velocity in m/s, V = accelerating voltage, z = charge

Therefore, ions with low m/z travel more rapidly than ions with high m/z . The time required for ions to travel through the field-free region can be accurately measured and used to calculate their m/z . The acceleration voltage and the length of the flight tube are two critical factors that affect the resolution and sensitivity of the measurement (Kinter & Sherman, 2000). To avoid increasing the length of the flight tube, a reflector is used. Current instruments have two reflectors (not shown in Figure 12) to accommodate “W-mode” ion travel, which improves resolution. In our study, we sought to maximise the sensitivity of the instrument, and therefore operated it in “V-mode”. Despite this, isotopic resolution was still achieved.

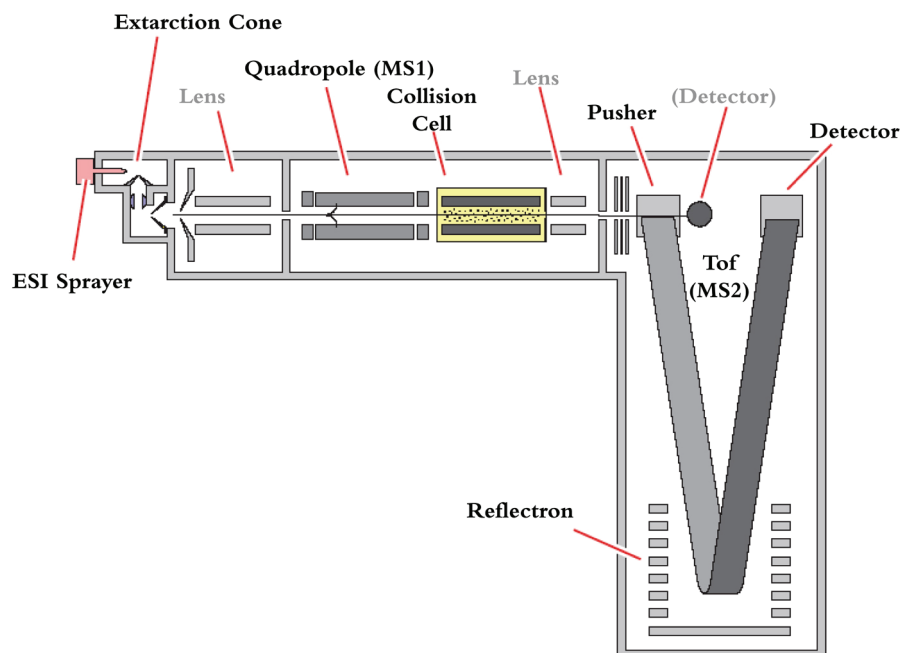


Figure 11. Schematic diagram of a hybrid Q-ToF mass spectrometer (V-mode). In a peptide fragmentation tandem mass spectrometry (MS/MS) experiment, the first stage involves selection of the desired precursor (intact) peptide ion, which is accomplished using the quadrupole mass filter (MS1). The second stage involves fragmentation in the collision cell, a hexapole lens system in which ions are accelerated in argon gas and all fragments are transmitted further without regard to their m/z . The third stage employs an orthogonal time-of-flight mass analyzer (MS2) which performs the mass analysis. For molecular weight measurement (MS), the first quadrupole is used in rf-only mode and transmits all peptide ions; no fragmentation is induced in the collision cell, and mass analysis is performed using the time-of-flight mass analyzer.

3.7 Peptide fragmentation

In mass spectrometric sequencing, the information that describes the amino acid sequence of the peptide is contained in the product ion spectrum. The product ion spectrum is acquired from a tandem mass spectrometry experiment by using low-energy collision-induced dissociation (CID). By using the first mass filter to select intact peptides, individual peptides within a narrow mass range can be accelerated in a collision cell. In the collision cell, which is typically filled with argon gas, intermolecular collisions convert kinetic energy (10eV to 50eV) to vibrational energy in the peptide ions. The ions' internal vibrational energy causes fragmentation according to

the mobile proton hypothesis. In the gas-phase, protons bind strongly to basic groups and remain bound even after collisions. In contrast, protons bound to less basic functional groups can move around 'within' the peptide to any one of its amide bonds by means of internal solvation. The migration of the mobile protons produces sub-populations of peptide ions and directs subsequent charge-site-directed fragmentation. Fragmentations occur primarily at the protonated amide bonds, producing the characteristic y and b-type ions, from the C- and N-termini respectively (Figure 12). Detection of the fragment ions in a second high resolution mass analyzer allows the fragments' masses to be measured with sufficient accuracy to determine the identities of their constituent amino acids. However, current instruments cannot distinguish between the isobaric amino acids isoleucine and leucine. Peptides produced by tryptic digestion are ideally suited for ESI and CID fragmentation. The charged C-terminal lysine and arginine residues of the fragments produced by tryptic cleavage give rise to predictable fragmentation spectra that can readily be related to specific amino acid sequences.

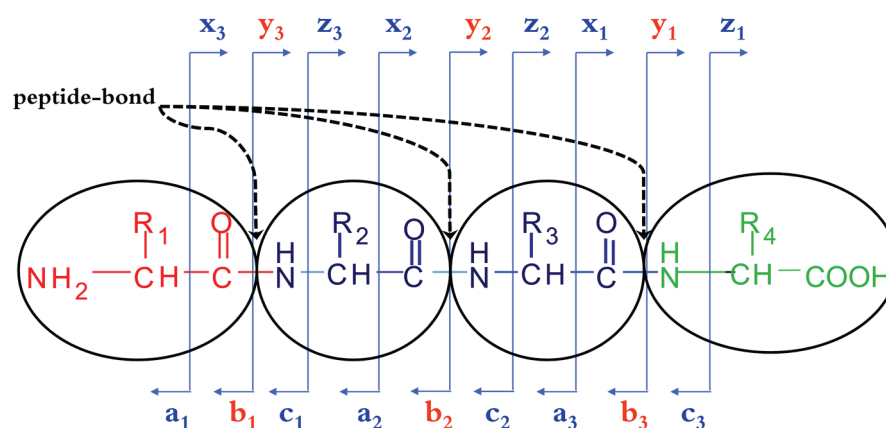


Figure 12. Peptide ion fragmentation nomenclature. Low-energy collision-induced dissociation primarily causes fragmentation at the peptide bonds (peptide or amide linkage) generating b- and y-ions. Internal fragmentations generate combinations of a-/x-ions or c-/z-ions as well as immonium ions characteristic of each amino acid.

3.8 Bioinformatics

3.8.1 Preprocessing of raw data

Continuous mass spectral data from the peptide ion (MS) and fragment ions (MS/MS) require processing before they can be used in database searches. The typical workflow for processing spectral data involves background reduction, smoothing, centroiding, off-set calibration, deisotoping and charge state deconvolution. The optimal processing settings for a given spectrum depend on the sample quality and the acquisition parameters used on the instrument. Different database search algorithms require the use of specific settings for optimal performance. Processing ultimately generates a basic list (peak list) of singly charged, monoisotopic masses for the intact peptide and its fragments.

3.8.2 Databases and search engines

The tools used for comparing mass spectrometry results to data in sequence databases are known as database-dependent algorithms because they are heavily dependent on the sequence database. Peak lists from processed raw data can be submitted to a large number of different commercial and publicly available database search engines (Matthiesen & Jensen, 2008). No matter what search engine is used, no results will be generated if the corresponding protein sequence is not present in the database. With common search parameters, exact amino acid matches are required. Good sequence coverage is available for arabidopsis because of the existence of extensive EST collections (from different ecotypes) and genomic sequencing data. Furthermore, there is a well-organized gene annotation system for arabidopsis, which allows for quick access to and use of genes with alternative splicing. Poplar is a much more difficult species to study, in part because the ecotype whose genome was sequenced is *P. trichocarpa* (Tuskan *et al.*, 2006), but research is often performed on hybrids or on natural populations of aspen (*P. tremula*) which exhibit substantial genetic variation (Luquez *et al.*, 2008). Furthermore, the whole-genome duplication event that occurred in poplar after its lineage diverged from that of arabidopsis greatly increased the complexity of its proteome. There are approximately 45 000 putative protein-coding genes in poplar, including nearly 8 000 pairs of paralogous genes (Tuskan *et al.*, 2006) whose protein products are too similar to be readily distinguished by MS/MS sequencing. The presence of paralogous genes and genetic variation makes the process of merging MS/MS results with data from a different source like an EST database unfeasible. Search engine programs like PLGS from Waters make it possible

to conduct a secondary database search to find modified peptides derived from proteins found in the preliminary first search. This facilitates the detection of mutations, genetic variations that affect codon usage and splicing, and post-translational protein modifications. Alternatively, one can use the *de novo* sequencing method, which does not rely on sequence databases. The process of algorithm-based *de novo* sequencing starts with the identification of significant ions, followed by examining the internal immonium ions to determine the peptide's amino acid composition, creation of b- and y-ion-evidence lists, determination of peptide sequences from the spectrum, generation of complete sequences (possible candidate sequences), and finally, scoring and ranking of the various calculated sequences (Reinders *et al.*, 2004). Manual or automated *de novo* sequencing is essential when evaluating peptides that have an ambiguous fragmentation pattern. While search engines will always identify the structure in the database that most closely matches the experimental data, the use of incomplete databases means that this method has the potential to generate false positive matches. Methods for estimating the false positive rate are therefore required.

Extracting peptide markers from LC-MS experiments and connecting these markers to individual proteins necessitates the use of rather elaborate data preparation techniques. The general workflow for the process is shown in Figure 13.

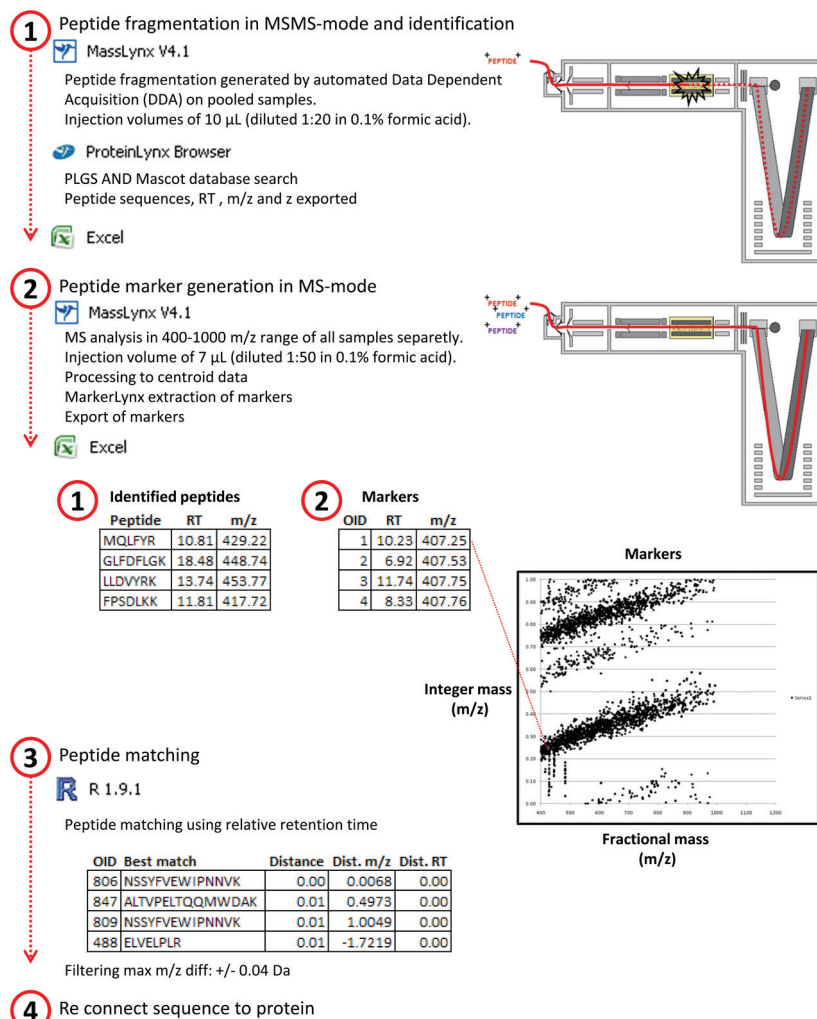


Figure 13. Workflow describing the extraction of markers and their use in identifying peptides

Careful experimental planning is required if one is to correctly interpret the results obtained from the analysed samples and their associated biological annotations. Where possible, we employed design of experiments (DoE) (Box *et al.*, 1978) in order to avoid drawing inaccurate conclusions from our experimental results. In the study described in paper II, a large set of samples (90 samples, with 10 biological replicates) was subjected preliminary metabolic profiling using GC-MS. The resulting metabolite profiles were subsequently explored using Principal Component Analysis (PCA) (Jolliffe,

2002; Wold *et al.*, 1987) in order to identify similarities and differences between the biological replicates, using the first principal component (along which the greatest variance in the data lies). The four most diverse biological samples for each genotype and internode were selected as candidates for further profiling in terms of their transcript and protein contents. The essential principle of this selection strategy was to ensure that the natural variability in the biological samples was not underestimated.

3.8.3 Bidirectional multivariate regression

The study described in paper II employed a bidirectional multivariate regression method (Trygg & Wold, 2003; Trygg, 2002) called O2PLS that identifies joint covariation between two data sets as well as systematic variation that is unique to each data set (Figure 14). Both the jointly covarying and unique sources of variation are composed of smaller units which are referred to as latent variables that describe independent effects in the data.

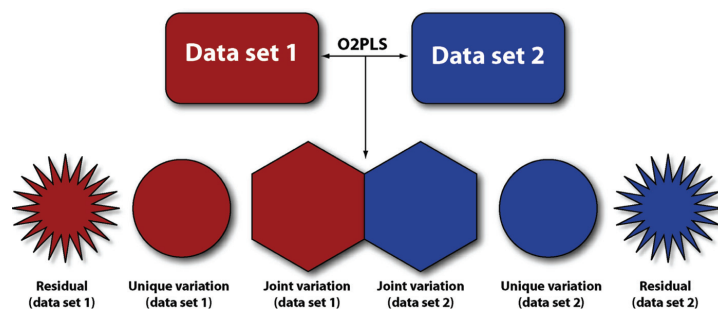


Figure 14. Overview of the components of the O2PLS model. The variation in each data set (matrix) is separated into predictive (joint) variation, data set specific variation and nonsystematic (residual) variation.

To handle multiple datasets, further processing is required. In paper II we integrate data from three different analytical platforms (transcript, protein and metabolite). Stepwise processing was employed to identify joint covariance, extract platform specific variation and finally, to analyse the matrices (Figure 15).

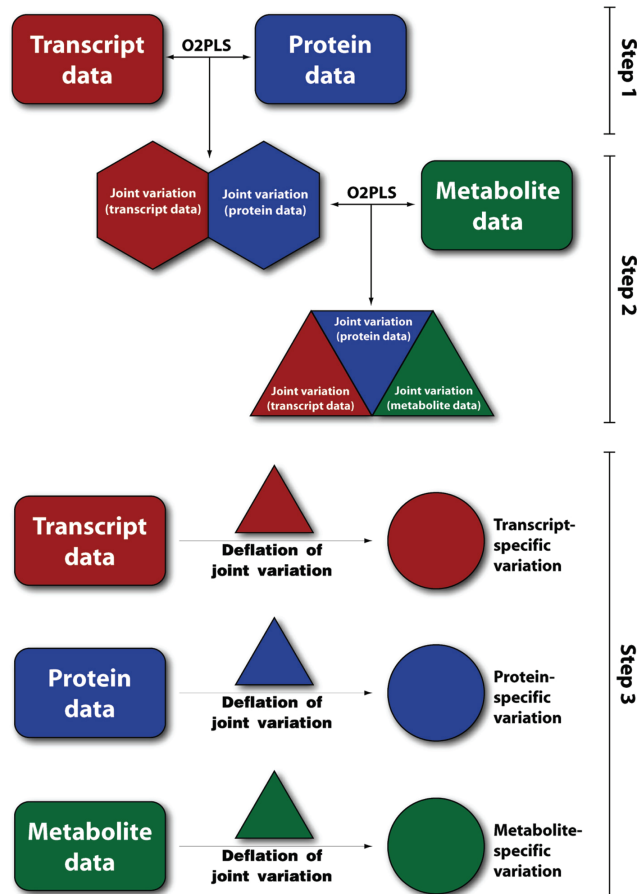


Figure 15. The O2PLS-based integration framework. The different steps used to identify joint and platform-specific variation for three data sets by means of O2PLS are shown. In Step 1, the joint covariance structures from the transcript and protein data sets are identified. In Step 2, the joint covariance structures from the joint transcript-protein variation in Step 1 and the metabolite data are utilized to identify the joint covariance between all data sets. In Step 3, the joint covariance structures are removed from each data set and specific systematic variation is extracted.

3.8.4 Visualizing bioinformatics data

We have used various approaches to obtain overviews of complex datasets and to capture the underlying biological relationships. To visualize the relationship between the genes of interest in paper II, we employed a multilayer network algorithm by forming the union of multiple minimal spanning trees (MST), each of which was constructed by random resamplings of a large data set of 1024 microarrays (Grönlund *et al.*, 2008). Such network visualization techniques can reveal interesting pathways by

application of the “guilt by association” principle. We integrated the network clusters with the correlation loading scores obtained from the O2PLS analysis to identify affected pathways.

4 Results and discussion

4.1 Composition of the plant transcriptional regulator

Mediator was originally identified as a central transcriptional coregulator in yeast (Kim *et al.*, 1994; Flanagan *et al.*, 1991; Kelleher *et al.*, 1990), but it is now clear that it also plays the same important role in higher eukaryotes (Bourbon *et al.*, 2004). A Mediator-like complex would be expected to play a major role in the regulation of wood formation at the transcriptional level during the appropriate stages in development. It was therefore puzzling that prior to our studies, no Mediator-like complex had been reported in plants. The nature of the Mediator complex we isolated from arabidopsis (paper I) suggests an explanation for this mystery: the subunits of the plant Mediator show very little sequence homology with the corresponding subunits from yeast and metazoans (Table 2).

Table 2. Homology between from arabidopsis and *S. cerevisiae* Mediator subunit 6, 7, 10, 14, 15 and 21 summarized

Gene	identity	similarity	length (aa)
AtMed6	24%	41%	191
AtMed7	33%	48%	188
AtMed10	26%	50%	139
AtMed14	21%	41%	171
AtMed15	26%	38%	132
AtMed21	24%	41%	139

This low homology is not so surprising considering that the arabidopsis genome encodes ~1 700 transcription factors, of which ~700 are specific to plants (Riechmann & Ratcliffe, 2000). Since these plant-specific

transcription factors will execute their function through Mediator, it was perhaps to be expected that the plant Mediator subunits would differ from those in yeast and humans.

In total, we identified 24 arabidopsis Mediator subunits by LC-ESI-MS/MS after purification by immunoprecipitation using antibodies raised against AtMed6, AtMed7, and ultimately, AtMed32 which would be named AtMed2 (Bourbon, 2008). We were unable to find arabidopsis homologs of the conserved Med1 subunit or the metazoan-specific Med26 subunit. The positions of gel spots (from 2D gels) and slices (from 1D gels) containing peptides that correspond to the identified Mediator subunits can be found in Figure 2 of paper I. Multiple spots corresponding to AtMed8 were observed. This is probably due to degradation, but may be indicative of post-translational modifications that affect its isoelectric focusing (Figure 16).

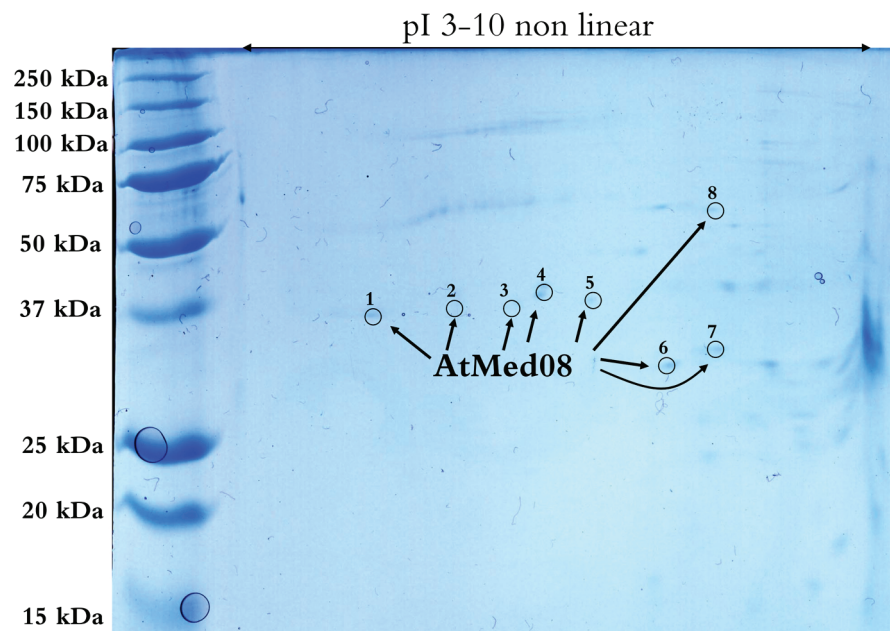


Figure 16. 2D gel electrophoresis of arabidopsis Mediator. Immunoprecipitated proteins were separated along the first dimension by isoelectric focusing followed by separation along the second dimension with SDS-PAGE (12%). Circles indicate spots where peptides belonging to AtMed8 were identified after in-gel tryptic digestion and LC-MS/MS analysis.

It is apparent that arabidopsis Mediator contains homologs of nearly all of the subunits from the head and middle domains in yeast, where, these subunits are situated in close proximity to RNAP II. However, it was not possible to identify homologs of many of the subunits that comprise the tail

domain in the yeast Mediator. After the publication of paper I, a comprehensive bioinformatics analysis of the evolutionary origin of the Mediator was undertaken; our recent results were examined along with the rest of the literature data (Bourbon, 2008). This detailed comparison of Mediator subunits from 70 eukaryotes with completed genome sequences (>90%) clarified the relationships between the divergent tail subunits. It was concluded that AtMed32 is a homolog of yMed2, AtMed33a/b is a homolog of yMed5a/b, and AtMed27 is a homolog of yMed3. Additionally, At5g63480 was found to be a homolog of yMed30; at the time of writing paper I, we had not been able to determine which Mediator domain it belonged to. The updated names of the Mediator subunits identified in our study are shown in Table 3. Of the four domains, the tail is the least conserved between different species. It is therefore likely that the plant-specific functions of the arabidopsis Mediator are carried out by the tail subunits.

Table 3. *Mediator subunits of arabidopsis identified by LC-ESI-MS/MS*

Protein name	New def. Bourbon	Accession number	Mediator module
Med32	Med2	At1g11760	Tail
Med27	Med3	At3g09180	Not identified
Med4	Med4	At5g02850	Middle
Med33a	Med5#1	At3g23590	Tail
Med33b	Med5#2	At2g48110	Tail
Med6	Med6	At3g21350	Head
Med7a	Med7#1	At5g03220	Middle
Med7b	Med7#2	At5g03500	Middle
Med8	Med8	At2g03070	Head
Med9	Med9	At1g55080	Middle
Med10a	Med10#1	At5g41910	Middle
Med10b	Med10#2	At1g26665	Middle
Med11	Med11	At3g01435	Head
Med14	Med14	At3g04740	Tail
Med15	Med15#1	At1g15780	Tail
Med16	Med16	At4g04920	Tail
Med17	Med17	At5g20170	Head
Med18	Med18	At2g22370	Head
Med19	Med19#1	At5g12230	Head
Med20	Med20#2	At2g28230	Head
Med21	Med21	At4g04780	Middle

Med22a	Med22#1	At1g16430	Head
Med22b	Med22#2	At1g07950	Head
Med23	Med23	At1g23230	Not identified
Med25	Med25	At1g25540	Not identified
Med28	Med28	At3g52860	Not identified
Med30	Med30	At5g63480	Not identified
Med31	Med31	At5g19910	Middle

The identification of At2g48110, which encodes AtMed5#2, (AtMed33b) as a subunit of Mediator is the first direct link between Mediator and the wood formation pathway. Point mutations at this locus have been shown to affect the phenylpropanoid pathway and lignification, (Stout *et al.*, 2008) which strongly suggests that Mediator is an important target of this pathway. Regulation is most likely effected by interactions between Mediator and the various regulatory TFs that are involved in the transcriptional networks that coordinate wood formation (Zhong & Ye, 2007). A member of the MYB TF family has already been shown to interact with a Mediator tail subunit in yeast two-hybrid screens (submitted). Paper I focuses on the AtMed25 subunit (PHYTOCHROME AND FLOWERING TIME 1, PFT1) and its involvement in flowering. It has recently been shown that the induction of flowering triggers xylem expansion in the hypocotyl and secondary growth in the roots in arabidopsis (Sibout *et al.*, 2008). The data support the hypothesis that flowering-related signals from the shoot control xylem expansion in the hypocotyl and roots.

The partitioning of the Mediator complex into core and peripheral elements is echoed in the structure of one of its subunits, AtMed25. We have shown that the N-terminal domain of AtMed25, which interacts with the Mediator core in human cells, is conserved in arabidopsis. However, human Med25 has a domain close to the C-terminus that interacts with the transcriptional activator VP16. This domain is conserved in metazoans but is not conserved in arabidopsis. By contrast, the corresponding C-terminal region is conserved between different plants (Figure 17). It seems that Mediator may have evolved to accommodate species-specific transcriptional regulators by incorporating subunits that possess one highly conserved domain that interacts with the other Mediator subunits together with a second, species-specific domain that interacts with the appropriate regulators of transcription.

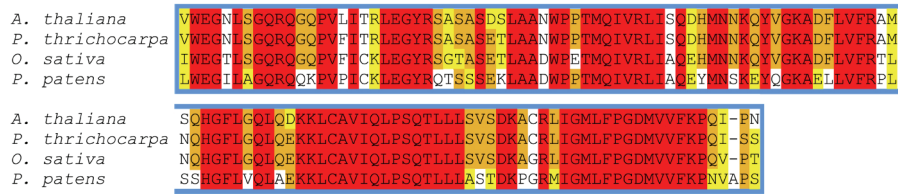


Figure 17. Partial sequence alignment of Med25 from the plants arabidopsis (*A. thaliana*), poplar (*P. trichocarpa*), *O. sativa*, and *P. patens*. The blue box indicate the region in the plant Mediator sequences that correspond to the minimal VP16 interaction domain of human Med25.

Interestingly, we identified peptides that map to different paralogs of the AtMed7, AtMed10, AtMed19, AtMed20, and AtMed22 subunits in our purified arabidopsis Mediator. Gene duplications that generated paralogs have played important roles in evolution. Some excellent examples of subfunctionalization and neofunctionalization after gene duplication were identified in arabidopsis by analysing 280 phylogenetically identified paralogous pairs. The expression of these pairs was measured in six organs; 85% of them exhibited significant gene by organ interactions, indicating that they have undergone subfunctionalization and/or neofunctionalization (Duarte *et al.*, 2006). Because we purified the arabidopsis Mediator from a cell suspension culture, we are unable to determine whether different paralogs are present in a single complex or even whether they are normally expressed in the same cell or cell type. It is possible that Mediator complexes containing different paralogs have different functions and that different paralogs are expressed in different tissues or in different developmental stages. Mediator has been reported to have tissue- and cell-specific functions. One way in which this could arise would be if the Mediator complexes of the different cell types or tissues contain different subunits. The corresponding paralogous genes in poplar may be involved in the regulation of tree specific traits like wood formation, long-term perennial growth, and seasonality.

Our studies on the isolation of Mediator also highlighted the importance of using an appropriate protein database in work of this kind. In this case, the NCBI database proved to have the greatest sequence coverage of At2g48110. Alternative gene models for the At2g48110 locus from the NCBI database were found to match the acetylated N-terminal peptide and a peptide having a slightly different C-terminus that was not present in the TAIR database. Moreover, we identified the full length mRNA arising from this locus; its size is consistent with a protein having a molecular weight of 16 kDa.

4.2 Correlation between transcripts, proteins and metabolites

Paper II describes a functional genomics strategy for the integrated analysis of transcript- (microarray), protein- and metabolite-data. The strategy relies on the use of designed experiments that are specifically tailored to the study at hand. However, the methodology is unsupervised in the sense that no prior knowledge of the sample is used in the modelling. We decided to demonstrate this strategy using poplar mutants that exhibit abnormal wood development. The mutants incorporate antisense constructs of PttMYB090 (PttMYB21a); this protein is involved in lignin biosynthesis (Karpinska *et al.*, 2004) and is part of the regulatory TF network that controls wood formation (Zhong *et al.*, 2009; Zhong *et al.*, 2008). We collected xylem tissue from wild-type (WT) and two transgenic lines (G3 and G5) at three internode positions (denoted A-C). Our experimental design (Figure 6) allowed us to analyse the growth characteristics and genotype effects individually and to look at synergistic effects. Multiple O2PLS models were used to integrate and reduce the large datasets generated. Two main effects were revealed by analysis of the joint covariation in the datasets; the effects were visible in all three profiles (transcript, protein, and metabolite). The first joint effect is the internode gradient, which reflects the common developmental progress of the samples, regardless of their genotype (Figure 18, A). The second joint effect reflects differences between the G5 and G3 genotypes, and is not sensitive to the precise internode examined (Figure 18, B). We refer to the first phenomenon as the internode effect and to the second as the genotype effect.

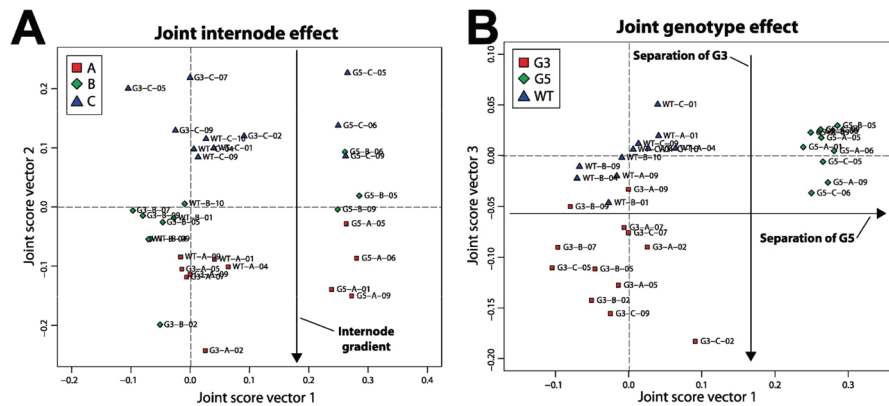


Figure 18. The internode and genotype effects. (A) The internode gradient is seen along the second joint score vector. (B) The genotype effect is shown as a combination of the first (G5 versus G3 and WT) and third joint score vectors (G3 versus G5 and WT).

These effects reflect the experimental setup shown in Figure 6. This verifies the fitness of the integrative methodology, as information on the nature of the samples was not factored into the experimental design. The results indicate that a considerable part (40% on average) of the existing variation in the steady-state system can be linked to changes in transcript levels, which are in turn reflected in the quantities of specific proteins observed and in the levels of individual metabolites. Much of the variation in the data reflects variation within individual profiles (e.g. within transcript levels) rather than interaction effects, and seems to be related to various housekeeping-type events. For example, we observed transcript-specific variation that was associated with processes involving chromatin remodelling and the assembly of protein-DNA complexes (Fig 8, paper II). This may be due in part to the fact that while microarray technology permits near-exhaustive monitoring of the transcriptome, the techniques used for proteomics and metabolomics generally measure only the most abundant proteins and metabolites. Secondly, processes like chromatin remodelling are regulated by complex post-transcriptional modifications and/or post-translational events that are not readily captured by the latter methods without using advanced enrichment techniques. This limitation restricts the scope for the simultaneous analysis of different kinds of -omics datasets using a global profiling approach; further technological advances are required to facilitate more comprehensive analyses.

The difficulties associated with combining transcript and proteomics data were highlighted in our studies of monolignol biosynthesis. We were able to identify the key enzymes involved in the biosynthetic pathway (Figure 19) but were only able to relate a few of these to specific extracted peptide markers. Additionally, we sought to identify the genes corresponding to these proteins using the microarray. However, the inherent complexity of the poplar genome, with its large number of duplicated genes, made it difficult to perform comparisons between the datasets. Even if it were possible to uniquely relate a peptide to a specific gene model, it would not be possible to uniquely identify the corresponding spot on the array. Each gene model on the spotted array usually has multiple reporters (PU-numbers) and it is essential to treat these in a uniform manner. Extensive filtering of the dataset was necessary, which made it difficult to obtain deeper insights into the biology underpinning the data.

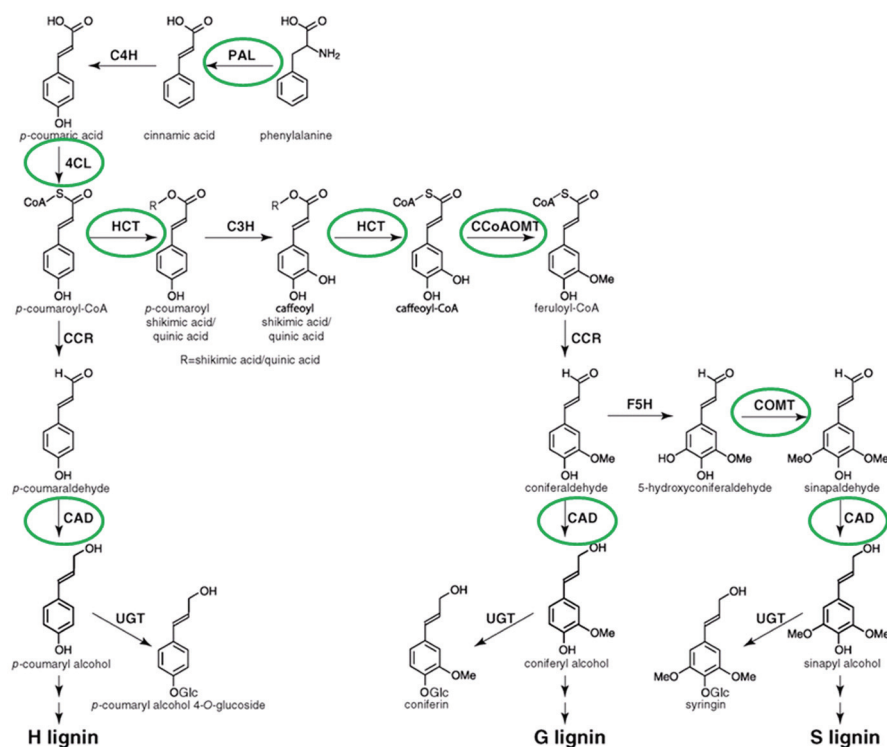


Figure 19. Overview of the enzymes detected when studying the biosynthesis of lignin. Enzymes were identified by extracting the highly water-soluble proteins from xylem tissue followed by LC MS/MS. The identified enzymes are indicated by green circles.

Despite the difficulties associated with using correlations between the transcript data and the proteomics data to identify unique gene models, the integrated analysis did reveal some interesting correlations. We were able to confirm previously-identified links between specific transcripts, proteins, and metabolites, and also discovered a number of potential new associations. The changes in the levels of photosynthesis-related transcripts, glucose metabolism proteins, and amino acids on going upwards from internode to internode (i.e. upwards along the developmental gradient) were as expected. Such changes are an unavoidable necessity of plant growth, and were accurately detected by all three profiling techniques. Specific effects due to the down-regulation of the PttMYB090 (PttMYB21a) gene, which is known to affect plant growth and lignin biosynthesis, were observed in the mutants (Karpinska *et al.*, 2004). We found that the normal growth gradient and the differences between the G5 and G3 genotypes are independent of one another. This would imply that with some exceptions, the mutants share the essential growth characteristics of a normal plant, although their

growth is somewhat slowed. The levels of important factors in both cell growth (specifically, tubulin and protein translation elongation factors) and lignin biosynthesis (COMT2 and CCoAOMT2) are heavily affected in both the transcript and protein data sets for the stronger G5 mutant. Importantly, it was observed that the reduction in transcript levels due to the G5 genotype resulted in an increase in the levels of the proteins encoded by those transcripts. Similar phenomena have previously been observed in yeast (Gygi *et al.*, 1999); they indicate that there are various active mechanisms of post-translational regulation and consequently, the levels of proteins will not be strongly correlated with those of the corresponding transcripts. Additionally, the changes in the levels of COMT2 and CCoAOMT1/2 in the G3 and G5 mutants correlated with changes in the levels of metabolites such as quinic acid that are involved in lignin biosynthesis.

4.3 Components of wood formation in the plasma membrane

In the work described in paper III, we identified more than 900 proteins in the plasma membranes isolated from leaves, xylem, and cambium/phloem obtained from young poplar trees (see Table 1, paper III and Supplemental Table 3, which is available online). One beneficial effect of using big trees in proteomics studies is that one can readily obtain rather large samples; we took advantage of this. A comprehensive protocol for the isolation and purification of plasma membranes was employed (Figure 9), involving the aqueous polymer two-phase partitioning technique previously described by Larsson *et al.* (1994). In the work described in paper I, we were able to successfully identify proteins digested after purification by 1D gel separation; 2D separation was not necessary. We therefore adopted the same approach when studying the plasma membrane proteins. 1D gel separation does not separate protein isoforms or other modified versions and provides higher protein to gel ratios than does 2D separation; this is favourable for gel digestion (Havlis *et al.*, 2003) and subsequent identification of the digested proteins. The protein bands were well focused, making it possible to use multiple injections, and individual bands was analysed over several mass ranges to increase the number of identified proteins. Because of the large number of proteins identified, we focused on proteins with transmembrane domains, as was done in previous studies. It should be noted that this restriction will exclude both true peripheral proteins that associate with the membrane by means of lipid anchors and also those that bind to the membrane as subunits of larger protein complexes. Up to 7% of the

identified proteins that lack a transmembrane domain are predicted to have lipid anchors.

The primary conclusion of this study was that more than 40% of the 956 proteins identified were found in the plasma membranes of all three tissues examined, and may thus be classified as “housekeeping” proteins. The most valuable results in terms of increasing our understanding of wood formation in poplar were obtained by comparing the differences in protein levels between the tissues. A graphical model of the plasma membrane and the complexes involved in cell wall synthesis was created, and cell wall related genes from arabidopsis were superimposed on this (Figure 20). Extra symbols were added to indicate proteins identified in a particular tissue. In several cases, we were able to identify unique peptides corresponding to specific poplar gene models, and thus to confirm their existence and location.

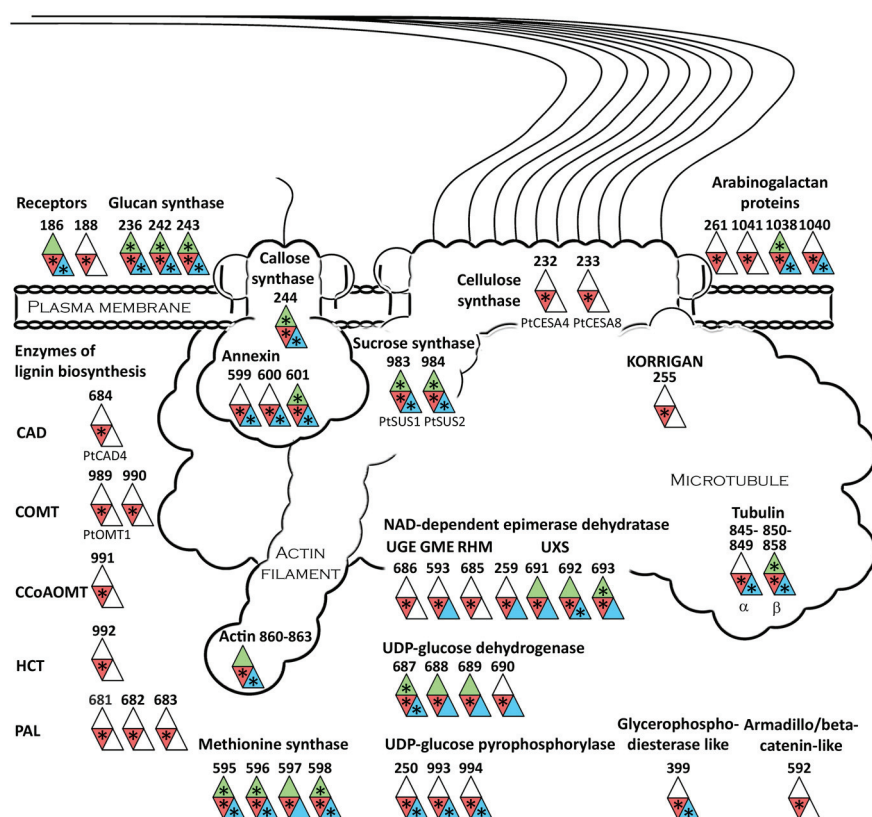


Figure 20. Proteins involved in wood formation. The figure shows a schematic model of the cellulose-synthesizing complex and other proteins associated with wood formation, such as

the enzymes involved in lignin biosynthesis. The numbers outside each symbol refer to the gene model ID in Table I of paper III for integral proteins and supplemental Table 3 for soluble proteins. Only proteins detected in the xylem plasma membranes are shown; they are denoted by a red triangle. Proteins that are also found in the plasma membranes of leaves are marked by a green triangle, while those that are found in cambium/phloem bear a blue triangle. A star in a colored field indicates that the protein is highly ranked within that tissue. Where available, the proteins' names in poplar are shown below the symbol. Abbreviations for the NAD-dependent epimerase dehydratase family are: UGE, UDP-D-glucose 4-epimerase; GME, GDP-D-mannose 3,5-epimerase; RHM, UDP-L-rhamnose synthase; UXS, UDP-D-apiose/xylose synthase. Abbreviations for the lignin enzymes are: CAD, cinnamyl alcohol dehydrogenase; COMT, caffeic acid O-methyltransferase; CCoAOMT, caffeoyl-CoA 3-O-methyltransferase; HCT, hydroxycinnamoyltransferase; PAL, phenylalanine ammonia-lyase.

8% of the 213 integral membrane proteins examined are involved in cell wall and carbohydrate metabolism. The other integral membrane proteins are discussed in detail in paper III; transport proteins constitute the largest class (41%), followed by receptors (14%), and proteins involved in membrane trafficking (8%). “Others,” of which various stress-induced proteins are the largest group, represent only 17%, and “unknowns” only 9%. Several of the ABC transporters show xylem-specific localization, indicating that they may be involved in secondary cell wall formation. Space constraints prohibited the inclusion of our phylogenetic analysis of the receptors in paper III. This analysis is shown in Figure 21. Some xylem-specific groups were identified from this important collection of signalling proteins. Of these proteins, only the LRR RLKs denoted by IDs 186 and 188 have (to the best of our knowledge) previously been associated with wood formation; this association was suggested on the basis of coexpression analyses (Persson *et al.*, 2005).

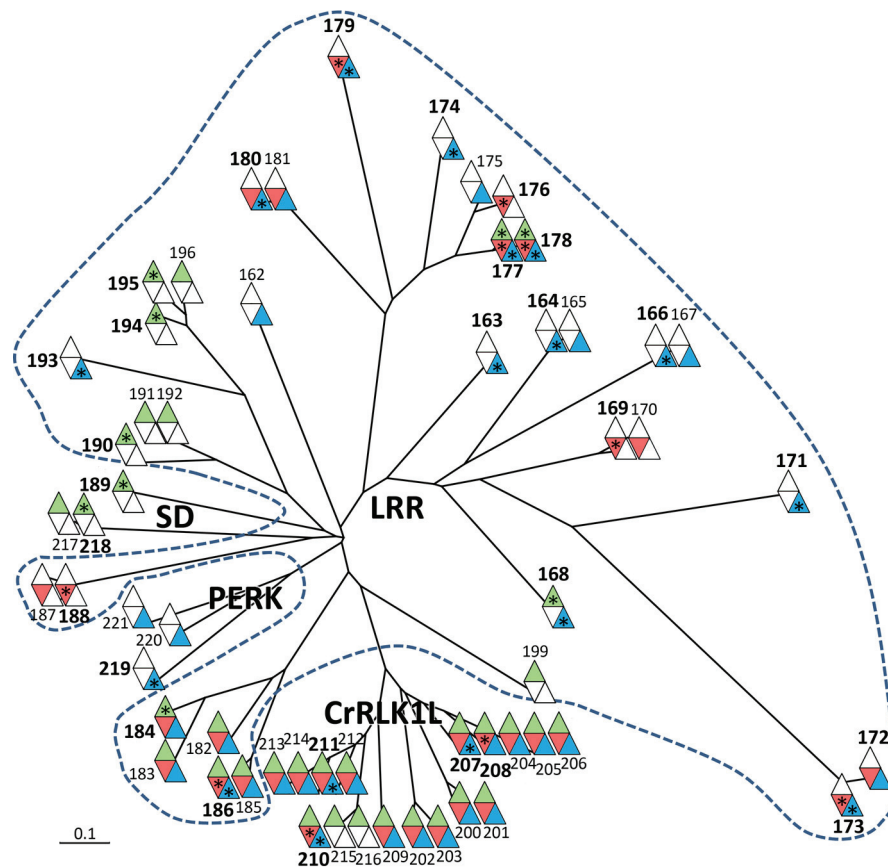


Figure 21. Phylogenetic Analysis of Plasma Membrane Receptor-Like Kinases (RLKs) and their Tissue Localization. Both highly-ranked and subset RLKs detected by mass spectrometry are included (two truncated sequences, denoted by IDs 197, 198 were excluded). The numbers outside each symbol refer to the gene model ID in Table 1, paper III (which contains top-ranked integral proteins only, indicated by numbers in bold) and Supplemental Table 2 (which contains both top rank and subset integral proteins). According to the phylogenetic analysis, four subfamilies of RLKs are represented: CrRLK1L (Catharantus roseus RLK1-Like), LRR (Leucine-Rich Repeat), PERK (Proline Extensin-like Receptor Kinase), and SD (S-locus glycoprotein-like Domain); the subfamilies are distinguished by their extracellular N-terminal ligand-binding domains (Shiu & Bleecker, 2001). Proteins found in leaves are denoted by a green triangle, those in xylem by a red triangle, and those in cambium/phloem by a blue triangle. A star in a colored field indicates that the protein is highly ranked within that tissue.

Proteins involved in cell wall and carbohydrate metabolism were most abundant in the xylem plasma membranes, in keeping with the role of the xylem in wood formation. Many peripheral proteins have essential roles in supporting the complex machinery that constructs the secondary cell wall.

We have identified proteins derived from a range of established genes and confirmed their expression and localization. Several novel candidates were also identified that have roles in less well-understood cellular processes that are nevertheless important in secondary cell wall formation.

5 Conclusions and future perspectives

The work described in this thesis has generated new insights into the molecular pathways leading to wood formation from a variety of perspectives, highlighting its universal importance in plant development.

We began by identifying the central transcriptional complex known as Mediator, which is involved in regulation of many essential processes in plants. We also identified subunits of the Mediator complex that are likely to be involved in regulating secondary cell wall formation. Subunits belonging to the tail of Mediator that are conserved within the plant kingdom are likely to perform plant specific functions. We demonstrated the existence of expressed and translated paralogous Mediator subunit genes that are unique to plants. A comparative analysis of Mediator genes from arabidopsis and poplar suggested that such paralogous genes may be involved in the regulation of tissue- and tree-specific traits such as wood formation. A previous study had shown that point mutations at the At2g48110 locus, which encodes Mediator subunit AtMed5#2 (AtMed33b), affect the phenylpropanoid pathway and lignification. This discovery may have practical applications: it has been stated that even modest decreases in the lignin content of biofuel feedstock brought about by such semidominant mutations could increase the efficiency of cellulosic biofuel production. Further, the reduced-lignin phenotypes caused by such transgenes may ultimately be more stable than those generated using RNAi-based strategies over successive generations or years in annual and perennial crops, respectively (Stout *et al.*, 2008). The AtMed25 Mediator subunit is also known as PHYTOCHROME AND FLOWERING TIME 1 (PFT1); as its name suggests, it is involved in flowering. It turns out that the induction of flowering also triggers xylem expansion in the hypocotyl and secondary growth in the roots of arabidopsis (Sibout *et al.*, 2008), suggesting that AtMed25 may have a role in wood formation. Unexpectedly we also found

a conserved domain in the plant protein AtMed25 that corresponds to the interaction domain of the human transcriptional activator Med25-VP16. The finding suggests that it may be possible to develop inducible chimeric systems for medical applications. Inducible systems derived from plant components can rely on chemicals that should have a negligible effect on mammals, or indeed, on animals in general (Corrado & Karali, 2009). From the perspective of plant research, considerable emphasis should be put on finding the activators that interact not only with Med25, but with all of the plant Mediator subunits. The identification of these interacting partners would make it possible to identify the events that trigger the transcriptional initiation of many known pathways back to their primary transcriptional initiation. The identification of such connections would in turn help in fleshing out the details of the mechanisms by which the pathways are regulated.

We suggest that Design of Experiments should be used in all future work involving the simultaneous analysis of levels of multiple different classes of biomolecules, as should our method for the integration of such data as described in paper II. Using these techniques, we were able to separate growth effect from genotype effects in mutants whose growth is severely affected. However, methods for combining data from transcriptomics, proteomics, metabolomics and other -omics are still in their infancy. In terms of proteomics in particular, many challenges remain to be overcome. While it is possible to achieve almost complete coverage of the transcriptome, proteomics is restricted to the analysis of relatively abundant species. Variations in the extent of transcription associated with nuclear processes performed by proteins of low abundance such as Mediator can be observed and isolated. However, although it is likely that such processes are also regulated at the protein level, regulation of this kind cannot be readily analysed with existing technology. In order to analyse more proteins, including sub populations of proteins that have undergone post-translational modification, it will be necessary to combine several different extraction procedures. In poplar and other trees, gene duplications make it difficult to distinguish certain proteins from one-another. Progress in discriminating between the isobaric amino acids isoleucine and leucine would be useful in this respect. Additionally, it would be desirable to automate the identification of markers (i.e. fragments having a specific mass that elute at a known time) that represent individual peptides. Automation could be achieved in a variety of ways. It is theoretically possible to use computational methods to predict the identity of all of the fragments that would be produced by digestion of all of the predicted proteins (of which poplar has

approximately 45 000) of a given species. One could then calculate those peptides' masses and their properties that influence MS detection; the physiochemical properties of peptides are more predictable than those of large proteins. Furthermore, liquid chromatography elution times can also be accurately modelled (Krokhin *et al.*, 2004). In theory, then, it is possible to predict the expected MS ion current of the digested proteome down to the isotope patterns generated by individual peptides. The growing collection of peptides from arabidopsis and poplar that have been identified with a high degree of confidence can be used to identify "proteotypic peptides" that are represented in specific tissues or samples, reducing the complexity of the analysis. As predictions get more accurate, there will be less need to perform numerous identification experiments on actual samples, and therefore smaller quantities of the sampled material will be needed. Ultimately, new algorithms will make it possible to rapidly identify markers with a high degree of confidence. However, the management and analysis of complex samples will still require extensive separation. Methods that exploit the readily-predicted isoelectric points of peptides are attractive in this context. However, current methods based on the isoelectric point require the use of ampholytes. Unfortunately, most ampholytes are incompatible with MS analysis, and so some effort should be focused on the development of innovative ampholyte-free methods that exploit peptide autofocusing (Tomásaron *et al.*, 2007).

Finally, we examined cell wall synthesis in poplar plasma membranes. This work was initially intended to be a screening study to aid in the planning of future experiments based on the acquisition of quantitative time series to examine processes that occur in the plasma membranes and are important for plant development. However, the "screening" proved to be very successful and identified close to 1 000 proteins, along with their tissue distribution patterns. The sheer number of proteins identified made it possible to analyse the behaviour and activity of different pathways in specific tissues. The identities of many proteins that had been assigned on the basis of single peptide hits from known pathways were supported by their association with others involved the same pathway. We were also able to identify the components of the machinery of cell wall synthesis and the proteins that support it. In future, it may be possible to extend this work by using gel free methods like LCMSE (Silva *et al.*, 2006) to better quantify the proteins. However, the gel based system are useful when analysing plasma membranes, as the gel effectively removes the detergents and lipids that are inevitably present as contaminants in separated plasma membranes.

Compared to other species that have previously been used in proteomics research, poplar presents a number of unusual challenges. We have been able to overcome some of these, and intend to apply the knowledge and experience we have gained by doing so to the analysis of even more challenging species such as spruce and pine.

6 Sammanfattning

Trädstammens vedceller (xylemet) består huvudsakligen av sekundär cellvägg uppbyggd av ett nätverk av cellulosa, hemicellulosa och lignin. Veden bildas av dotterceller från ett tunnt lager av celler (kambiet) till xylemceller som i sin tur genomgår cellexpansion, sekundär cellväggssyntes, lignifiering och apoptos. Under hela utvecklingen av den komplicerade matrisen (veden) deltar ett stort antal proteiner. Biosyntesen av xylemet kräver därför en strikt reglering på molekylär nivå främst genom transkriptionell reglering med olika posttranskriptionella processer.

Studierna i denna avhandling baserar sig på vätskekromatografi kopplad till masspektrometri i syfte att kartlägga biosyntesen av ved. Denna teknik är för närvarande den dominerande för proteinanalys. I den första studien identifieras de proteiner i arabidopsis som ingår i den centrala transkriptionella enheten som kallas Mediatoren. Mediator-proteinkomplexet koordinerar de flesta essentiella processerna i växter. Specifika subenheter i komplexet, som genom genduplikation och evolution existerar i olika former, utgör troliga nav för sekundär cellväggsbildning. Punktmutationer i subenhet Med5#2 påverkar således lignifieringen i sekundärvägen. Dessutom kan xylemexpansionen indirekt kontrolleras av subenhet Med25. En strategi för att integrera transkript-, protein- och metabolitdata har utvecklats och prövats på mutanter av poppel som uppvisar förändrad vedbildning. Resultaten visade att ligninbiosyntesen var starkt påverkad både på transkript och protein nivå i mutanten med starkast fenotyp. Intressantast är att då yttrycket på vissa transkript sjönk, ökade korresponderande protein i vävnaden. Slutligen analyserades fronten av cellväggssyntesen i plasmamembran från poppel. Större delen av de proteiner som är väl dokumenterade cellväggsenzymer och associerade till cellväggssyntesen identifierades. Flertalet nya kandidater identifierades såväl receptorer, transportörer som strukturella proteiner.

References

- Anon. (2008). *Statistical Yearbook of Forestry 2008*. Jönköping: Swedish Forest Agency. ISBN 978-91-88462-78-7.
- Aspeborg, H., Schrader, J., Coutinho, P.M., Stam, M., Kallas, A., Djerbi, S., Nilsson, P., Denman, S., Amini, B., Sterky, F., Master, E., Sandberg, G., Mellerowicz, E., Sundberg, B., Henrissat, B. & Teeri, T.T. (2005). Carbohydrate-active enzymes involved in the secondary cell wall biogenesis in hybrid aspen. *Plant Physiology* 137(3), 983-997.
- Asturias, F.J., Jiang, Y.W., Myers, L.C., Gustafsson, C.M. & Kornberg, R.D. (1999). Conserved structures of mediator and RNA polymerase II holoenzyme. *Science* 283(5404), 985-987.
- Bailey-Serres, J., Sorenson, R. & Juntawong, P. (2009). Getting the message across: cytoplasmic ribonucleoprotein complexes. *Trends in Plant Science* 14(8), 443-453.
- Barbazuk, W.B., Fu, Y. & McGinnis, K.M. (2008). Genome-wide analyses of alternative splicing in plants: Opportunities and challenges. *Genome Research* 18(9), 1381-1392.
- Blanc, G., Barakat, A., Guyot, R., Cooke, R. & Delseny, I. (2000). Extensive duplication and reshuffling in the arabidopsis genome. *Plant Cell* 12(7), 1093-1101.
- Boerjan, W., Ralph, J. & Baucher, M. (2003). Lignin biosynthesis. *Annual Review of Plant Biology* 54, 519-546.
- Borchers, C., Peter, J.F., Hall, M.C., Kunkel, T.A. & Tomer, K.B. (2000). Identification of in-gel digested proteins by complementary peptide mass fingerprinting and tandem mass spectrometry data obtained on an electrospray ionization quadrupole time-of-flight mass spectrometer. *Analytical Chemistry* 72(6), 1163-1168.
- Bourbon, H.M. (2008). Comparative genomics supports a deep evolutionary origin for the large, four-module transcriptional mediator complex. *Nucleic Acids Research* 36(12), 3993-4008.
- Bourbon, H.M., Aguilera, A., Ansari, A.Z., Asturias, F.J., Berk, A.J., Bjorklund, S., Blackwell, T.K., Borggreffe, T., Carey, M., Carlson, M., Conaway, J.W., Conaway, R.C., Emmons, S.W., Fondell, J.D., Freedman, L.P., Fukasawa, T., Gustafsson, C.M., Han, M., He, X., Herman, P.K., Hinnebusch, A.G., Holmberg, S., Holstege, F.C., Jaehning, J.A., Kim, Y.J., Kuras, L., Leutz, A., Lis, J.T., Meisterernest, M., Naar, A.M., Nasmyth, K., Parvin, J.D., Ptashne, M., Reinberg, D., Ronne, H., Sadowski, I., Sakurai, H., Sipiczki, M., Sternberg, P.W., Stillman, R.

- D.J., Strich, R., Struhl, K., Svejstrup, J.Q., Tuck, S., Winston, F., Roeder, R.G. & Kornberg, R.D. (2004). A unified nomenclature for protein subunits of Mediator complexes linking transcriptional regulators to RNA polymerase II. *Molecular Cell* 14(5), 553-557.
- Box, G.E.P., Hunter, W.G. & Hunter, J.S. (1978). *Statistics for Experimenters: An Introduction to Design, Data Analysis, and Model Building*. New York: John Wiley & Sons, Inc. ISBN 0-471-09315-7.
- Brett, C.T. & Waldron, K.W. (1996). *Topics in Plant Functional Biology, 1. Physiology and biochemistry of plant cell walls, Second edition*: Chapman and Hall Ltd.; Chapman and Hall, Inc. (Topics in Plant Functional Biology, 1; Physiology and biochemistry of plant cell walls, Second edition. ISBN 0-412-58060-8.
- Carrari, F., Baxter, C., Usadel, B., Urbanczyk-Wochniak, E., Zanon, M.I., Nunes-Nesi, A., Nikiforova, V., Centro, D., Ratzka, A., Pauly, M., Sweetlove, L.J. & Fernie, A.R. (2006). Integrated analysis of metabolite and transcript levels reveals the metabolic shifts that underlie tomato fruit development and highlight regulatory aspects of metabolic network behavior. *Plant Physiology* 142(4), 1380-1396.
- Carthew, R.W. & Sontheimer, E.J. (2009). Origins and Mechanisms of miRNAs and siRNAs. *Cell* 136(4), 642-655.
- Cassab, G.I. & Varner, J.E. (1988). Cell-wall proteins. *Annual Review of Plant Physiology and Plant Molecular Biology* 39, 321-353.
- Clish, C.B., Davidov, E., Oresic, M., Plasterer, T.N., Lavine, G., Londo, T., Meys, M., Snell, P., Stochaj, W., Adourian, A., Zhang, X., Morel, N., Neumann, E., Verheij, E., Vogels, J.T., Havekes, L.M., Afeyan, N., Regnier, F., van der Greef, J. & Naylor, S. (2004). Integrative biological analysis of the APOE*3-leiden transgenic mouse. *OMICS* 8(1), 3-13.
- Corrado, G. & Karali, M. (2009). Inducible gene expression systems and plant biotechnology. *Biotechnology Advances* 27(6), 733-743.
- Crick, F. (1970). Central dogma of molecular biology. *Nature* 227(5258), 561-563.
- Davies, G.J. & Henrissat, B. (2002). Structural enzymology of carbohydrate-active enzymes: implications for the post-genomic era. *Biochemical Society Transactions* 30, 291-297.
- de Hoffmann, E. & Stroobant, V. (2001). *Mass Spectrometry: Principles and Applications*. 2. ed: John Wiley & Sons, Inc. ISBN 0-471-48566-7.
- Déjardin, A., Laurans, F., Arnaud, D., Breton, C., Pilate, G. & Leplé, J.-C. (2010). Wood formation in Angiosperms. *Comptes Rendus Biologies* 333(4), 325-334.
- Demura, T. & Fukuda, H. (2007). Transcriptional regulation in wood formation. *Trends in Plant Science* 12(2), 64-70.
- Donaldson, L., Hague, J. & Snell, R. (2001). Lignin distribution in coppice poplar, linseed and wheat straw. *Holzforschung* 55(4), 379-385.
- Duarte, J.M., Cui, L.Y., Wall, P.K., Zhang, Q., Zhang, X.H., Leebens-Mack, J., Ma, H., Altman, N. & dePamphilis, C.W. (2006). Expression pattern shifts following duplication indicative of subfunctionalization and neofunctionalization in regulatory genes of *Arabidopsis*. *Molecular Biology and Evolution* 23(2), 469-478.
- Farnham, P.J. (2009). Insights from genomic profiling of transcription factors. *Nature Reviews Genetics* 10(9), 605-616.

- Flanagan, P.M., Kelleher, R.J., Sayre, M.H., Tschochner, H. & Kornberg, R.D. (1991). A mediator required for activation of RNA polymerase II transcription *in vitro*. *Nature* 350(6317), 436–438.
- Geisler-Lee, J., Geisler, M., Coutinho, P.M., Segerman, B., Nishikubo, N., Takahashi, J., Aspeborg, H., Djerbi, S., Master, E., Andersson-Gunneras, S., Sundberg, B., Karpinski, S., Teeri, T.T., Kleczkowski, L.A., Henrissat, B. & Mellerowicz, E.J. (2006). Poplar carbohydrate-active enzymes. Gene identification and expression analyses. *Plant Physiology* 140(3), 946–962.
- Giavalisco, P., Nordhoff, E., Lehrach, H., Gobom, J. & Klose, J. (2003). Extraction of proteins from plant tissues for two-dimensional electrophoresis analysis. *Electrophoresis* 24(1–2), 207–216.
- Grönlund, A., Bhalerao, R.P. & Karlsson, J. (2008). Modular gene expression in Poplar: a multilayer network approach. *New Phytologist* 181(2), 315–322.
- Gygi, S.P., Rochon, Y., Franza, B.R. & Aebersold, R. (1999). Correlation between protein and mRNA abundance in yeast. *Molecular and Cellular Biology* 19(3), 1720–1730.
- Havlis, J., Thomas, H., Sebela, M. & Shevchenko, A. (2003). Fast-response proteomics by accelerated in-gel digestion of proteins. *Analytical Chemistry* 75(6), 1300–1306.
- Hirai, M.Y., Klein, M., Fujikawa, Y., Yano, M., Goodenowe, D.B., Yamazaki, Y., Kanaya, S., Nakamura, Y., Kitayama, M., Suzuki, H., Sakurai, N., Shibata, D., Tokuhisa, J., Reichelt, M., Gershenzon, J., Papenbrock, J. & Saito, K. (2005). Elucidation of gene-to-gene and metabolite-to-gene networks in *Arabidopsis* by integration of metabolomics and transcriptomics. *Journal of Biological Chemistry* 280(27), 25590–25595.
- Hirai, M.Y., Yano, M., Goodenowe, D.B., Kanaya, S., Kimura, T., Awazuhara, M., Arita, M., Fujiwara, T. & Saito, K. (2004). Integration of transcriptomics and metabolomics for understanding of global responses to nutritional stresses in *Arabidopsis thaliana*. *Proceedings of the National Academy of Sciences of the United States of America* 101(27), 10205–10210.
- Jansson, S. & Douglas, C.J. (2007). *Populus*: A model system for plant biology. *Annual Review of Plant Biology* 58, 435–458.
- Jensen, O.N. (2006). Interpreting the protein language using proteomics. *Nature Reviews Molecular Cell Biology* 7(6), 391–403.
- Jolliffe, I.T. (2002). *Principal Component Analysis*. 2. ed. New York: Springer. ISBN 978-0387954424.
- Karpinska, B., Karlsson, M., Srivastava, M., Stenberg, A., Schrader, J., Sterky, F., Bhalerao, R. & Wingsle, G. (2004). MYB transcription factors are differentially expressed and regulated during secondary vascular tissue development in hybrid aspen. *Plant Molecular Biology* 56(2), 255–270.
- Kelleher, R.J., Flanagan, P.M. & Kornberg, R.D. (1990). A novel mediator between activator proteins and the RNA polymerase II transcription apparatus. *Cell* 61(7), 1209–1215.
- Khidekel, N. & Hsieh-Wilson, L.C. (2004). A 'molecular switchboard' – covalent modifications to proteins and their impact on transcription. *Organic & Biomolecular Chemistry* 2(1), 1–7.

- Kim, Y.J., Bjorklund, S., Li, Y., Sayre, M.H. & Kornberg, R.D. (1994). A multiprotein mediator of transcriptional activation and its interaction with the C-terminal repeat domain of RNA polymerase II. *Cell* 77(4), 599-608.
- Kinter, M. & Sherman, N.E. (2000). *Protein Sequencing and Identification Using Tandem Mass Spectrometry*. New York: John Wiley & Sons, Inc. ISBN 978-0471322498.
- Kleno, T.G., Kiehr, B., Baunsgaard, D. & Sidelmann, U.G. (2004). Combination of 'omics' data to investigate the mechanism(s) of hydrazine-induced hepatotoxicity in rats and to identify potential biomarkers. *Biomarkers* 9(2), 116-138.
- Kolbe, A., Oliver, S.N., Fernie, A.R., Stitt, M., van Dongen, J.T. & Geigenberger, P. (2006). Combined transcript and metabolite profiling of *Arabidopsis* leaves reveals fundamental effects of the thiol-disulfide status on plant metabolism. *Plant Physiology* 141(2), 412-422.
- Koleske, A.J. & Young, R.A. (1994). An RNA polymerase II holoenzyme responsive to activators. *Nature* 368(6470), 466-469.
- Krokhin, O.V., Craig, R., Spicer, V., Ens, W., Standing, K.G., Beavis, R.C. & Wilkins, J.A. (2004). An improved model for prediction of retention times of tryptic peptides in ion pair reversed-phase HPLC - Its application to protein peptide mapping by off-line HPLC-MALDI MS. *Molecular & Cellular Proteomics* 3(9), 908-919.
- Kumar, M., Thammannagowda, S., Bulone, V., Chiang, V., Han, K.H., Joshi, C.P., Mansfield, S.D., Mellerowicz, E., Sundberg, B., Teeri, T. & Ellis, B.E. (2009). An update on the nomenclature for the cellulose synthase genes in *Populus*. *Trends in Plant Science* 14(5), 248-254.
- Larson, P.R. (1994). *The vascular cambium: development and structure*. Berlin Germany: Springer-Verlag. ISBN 3-540-57165-5.
- Larsson, C., Sommarin, M. & Widell, S. (1994). Isolation of highly purified plant plasma-membranes and separation of inside-out and right-side-out vesicles. *Aqueous Two-Phase Systems* 228, 451-469.
- Latchman, D.S. (1998). *Eukaryotic transcription factors*. 3. ed. San Diego: Academic Press. ISBN 978-0124371767.
- Lin, X.Y., Kaul, S.S., Rounsley, S., Shea, T.P., Benito, M.I., Town, C.D., Fujii, C.Y., Mason, T., Bowman, C.L., Barnstead, M., Feldblyum, T.V., Buell, C.R., Ketchum, K.A., Lee, J., Ronning, C.M., Koo, H.L., Moffat, K.S., Cronin, L.A., Shen, M., Pai, G., Van Aken, S., Umayam, L., Tallon, L.J., Gill, J.E., Adams, M.D., Carrera, A.J., Creasy, T.H., Goodman, H.M., Somerville, C.R., Copenhaver, G.P., Preuss, D., Nierman, W.C., White, O., Eisen, J.A., Salzberg, S.L., Fraser, C.M. & Venter, J.C. (1999). Sequence and analysis of chromosome 2 of the plant *Arabidopsis thaliana*. *Nature* 402(6763), 761-768.
- Luquez, V., Hall, D., Albrechtsen, B.R., Karlsson, J., Ingvarsson, P. & Jansson, S. (2008). Natural phenological variation in aspen (*Populus tremula*): the SwAsp collection. *Tree Genetics & Genomes* 4(2), 279-292.
- Malik, S. & Roeder, R.G. (2005). Dynamic regulation of pol II transcription by the mammalian Mediator complex. *Trends in Biochemical Sciences* 30(5), 256-263.
- Matthiesen, R. & Jensen, O.N. (2008). Analysis of mass spectrometry data in proteomics. In: Keith, J.M. (Ed.) *Methods in Molecular Biology: Structure, Function and Applications*.

pp. 105–122 Humana Press Inc. (Methods in Molecular Biology : 453; 2). ISBN 978-1-60327-428-9.

- Mayer, K., Schuller, C., Wambutt, R., Murphy, G., Volckaert, G., Pohl, T., Dusterhoft, A., Stiekema, W., Entian, K.D., Terry, N., Harris, B., Ansorge, W., Brandt, P., Grivell, L., Rieger, M., Weichselgartner, M., de Simone, V., Obermaier, B., Mache, R., Muller, M., Kreis, M., Delseny, M., Puigdomenech, P., Watson, M., Schmidheini, T., Reichert, B., Portatelle, D., Perez-Alonso, M., Boutry, M., Bancroft, I., Vos, P., Hoheisel, J., Zimmermann, W., Wedler, H., Ridley, P., Langham, S.A., McCullagh, B., Bilham, L., Robben, J., Van der Schueren, J., Grymonprez, B., Chuang, Y.J., Vandenbussche, F., Braeken, M., Weltjens, I., Voet, M., Bastiaens, I., Aert, R., Defoor, E., Weitzenegger, T., Bothe, G., Ramsperger, U., Hilbert, H., Braun, M., Holzer, E., Brandt, A., Peters, S., van Staveren, M., Dirkse, W., Mooijman, P., Lankhorst, R.K., Rose, M., Hauf, J., Kotter, P., Berneiser, S., Hempel, S., Feldpausch, M., Lamberth, S., Van den Daele, H., De Keyser, A., Buysschaert, C., Gielen, J., Villarroel, R., De Clercq, R., Van Montagu, M., Rogers, J., Cronin, A., Quail, M., Bray-Allen, S., Clark, L., Doggett, J., Hall, S., Kay, M., Lennard, N., McLay, K., Mayes, R., Pettett, A., Rajandream, M.A., Lyne, M., Benes, V., Rechmann, S., Borkova, D., Blocker, H., Scharfe, M., Grimm, M., Lohnert, T.H., Dose, S., de Haan, M., Maarse, A., Schafer, M., Muller-Auer, S., Gabel, C., Fuchs, M., Fartmann, B., Granderath, K., Dauner, D., Herzl, A., Neumann, S., Argiriou, A., Vitale, D., Liguori, R., Piravandi, E., Massenet, O., Quigley, F., Clabaud, G., Mundlein, A., Felber, R., Schnabl, S., Hiller, R., Schmidt, W., Lecharny, A., Aubourg, S., Chefdor, F., Cooke, R., Berger, C., Montfort, A., Casacuberta, E., Gibbons, T., Weber, N., Vandenbol, M., Bargues, M., Terol, J., Torres, A., Perez-Perez, A., Purnelle, B., Bent, E., Johnson, S., Tacon, D., Jesse, T., Heijnen, L., Schwarz, S., Scholler, P., Heber, S., Francs, P., Bielke, C., Frishman, D., Haase, D., Lemcke, K., Mewes, H.W., Stocker, S., Zaccaria, P., Bevan, M., Wilson, R.K., de la Bastide, M., Habermann, K., Parnell, L., Dedhia, N., Gnoj, L., Schutz, K., Huang, E., Spiegel, L., Sehkoni, M., Murray, J., Sheet, P., Cordes, M., Abu-Threideh, J., Stoneking, T., Kalicki, J., Graves, T., Harmon, G., Edwards, J., Latreille, P., Courtney, L., Cloud, J., Abbott, A., Scott, K., Johnson, D., Minx, P., Bentley, D., Fulton, B., Miller, N., Greco, T., Kemp, K., Kramer, J., Fulton, L., Mardis, E., Dante, M., Pepin, K., Hillier, L., Nelson, J., Spieth, J., Ryan, E., Andrews, S., Geisel, C., Layman, D., Du, H., Ali, J., Berghoff, A., Jones, K., Drone, K., Cotton, M., Joshi, C., Antoniou, B., Zidanic, M., Strong, C., Sun, H., Lamar, B., Yordan, C., Ma, P., Zhong, J., Preston, R., Vil, D., Shekher, M., Matero, A., Shah, R., Swaby, I., European Union Arabidopsis Genome; Washington Univ, P.E.B. & Arabidopsis, S. (1999). Sequence and analysis of chromosome 4 of the plant *Arabidopsis thaliana*. *Nature* 402(6763), 769–777.
- Mellerowicz, E.J., Baucher, M., Sundberg, B. & Boerjan, W. (2001). Unravelling cell wall formation in the woody dicot stem. *Plant Molecular Biology* 47(1–2), 239–274.
- Myer, V.E. & Young, R.A. (1998). RNA polymerase II holoenzymes and subcomplexes. *Journal of Biological Chemistry* 273(43), 27757–27760.

- Myers, L.C. & Kornberg, R.D. (2000). Mediator of transcriptional regulation. *Annual Review of Biochemistry* 69, 729-749.
- Oakley, R.V., Wang, Y.S., Ramakrishna, W., Harding, S.A. & Tsai, C.J. (2007). Differential expansion and expression of α - and β -tubulin gene families in *Populus*. *Plant Physiology* 145, 961-973.
- Olsen, J.V., Blagoev, B., Gnäd, F., Macek, B., Kumar, C., Mortensen, P. & Mann, M. (2006). Global, in vivo, and site-specific phosphorylation dynamics in signaling networks. *Cell* 127(3), 635-648.
- Onodera, Y., Haag, J.R., Ream, T., Nunes, P.C., Pontes, O. & Pikaard, C.S. (2005). Plant nuclear RNA polymerase IV mediates siRNA and DNA methylation-dependent heterochromatin formation. *Cell* 120(5), 613-622.
- Oresic, M., Clish, C.B., Davidov, E.J., Verheij, E., Vogels, J., Havekes, L.M., Neumann, E., Adourian, A., Naylor, S., van der Greef, J. & Plasterer, T. (2004). Phenotype characterisation using integrated gene transcript, protein and metabolite profiling. *Applied Bioinformatics* 3(4), 205-207.
- Pauly, M. & Keegstra, K. (2008). Cell-wall carbohydrates and their modification as a resource for biofuels. *Plant Journal* 54(4), 559-568.
- Persson, S., Wei, H.R., Milne, J., Page, G.P. & Somerville, C.R. (2005). Identification of genes required for cellulose synthesis by regression analysis of public microarray data sets. *Proceedings of the National Academy of Sciences of the United States of America* 102(24), 8633-8638.
- Pflüger, J. & Wagner, D. (2007). Histone modifications and dynamic regulation of genome accessibility in plants. *Current Opinion in Plant Biology* 10(6), 645-652.
- Ream, T.S., Haag, J.R., Wierzbicki, A.T., Nicora, C.D., Norbeck, A.D., Zhu, J.-K., Hagen, G., Guilfoyle, T.J., Pasa-Tolic, L. & Pikaard, C.S. (2009). Subunit Compositions of the RNA-Silencing Enzymes Pol IV and Pol V Reveal Their Origins as Specialized Forms of RNA Polymerase II. *Molecular Cell* 33(2), 192-203.
- Reddy, A.S.N. (2007). Alternative splicing of pre-messenger RNAs in plants in the genomic era. *Annual Review of Plant Biology* 58, 267-294.
- Reinders, J., Lewandrowski, U., Moebius, J., Wagner, Y. & Sickmann, A. (2004). Challenges in mass spectrometry-based proteomics. *Proteomics* 4(12), 3686-3703.
- Riechmann, J.L. & Ratcliffe, O.J. (2000). A genomic perspective on plant transcription factors. *Current Opinion in Plant Biology* 3(5), 423-434.
- Rischer, H., Oresic, M., Seppanen-Laakso, T., Katajamaa, M., Lammertyn, F., Ardiles-Diaz, W., Van Montagu, M.C., Inze, D., Oksman-Caldentey, K.M. & Goossens, A. (2006). Gene-to-metabolite networks for terpenoid indole alkaloid biosynthesis in *Catharanthus roseus* cells. *Proceedings of the National Academy of Sciences of the United States of America* 103(14), 5614-5619.
- Samuelson, C.O., Baraznenok, V., Khorosjutina, O., Spahr, H., Kieselbach, T., Holmberg, S. & Gustafsson, C.M. (2003). TRAP230/ARC240 and TRAP240/ARC250 Mediator subunits are functionally conserved through evolution. *Proceedings of the National Academy of Sciences of the United States of America* 100(11), 6422-6427.
- Schena, M., Shalon, D., Davis, R.W. & Brown, P.O. (1995). Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science* 270(5235), 467-470.

- Schmidt, J.T. (2008). *Functional Studies of Selected Extracellular Carbohydrate-Active Hydrolases in Wood Formation*. Diss. Umeå:SLU.
- Shi, R., Sun, Y.-H., Li, Q., Heber, S., Sederoff, R. & Chiang, V.L. (2010). Towards a systems approach for lignin biosynthesis in *Populus trichocarpa*: Transcript abundance and specificity of the monolignol biosynthetic genes. *Plant and Cell Physiology* 51(1), 144-163.
- Shiu, S.H. & Bleecker, A.B. (2001). Receptor-like kinases from *Arabidopsis* form a monophyletic gene family related to animal receptor kinases. *Proceedings of the National Academy of Sciences of the United States of America* 98(19), 10763-10768.
- Sibout, R., Plantegenet, S. & Hardtke, C.S. (2008). Flowering as a condition for xylem expansion in *Arabidopsis* hypocotyl and root. *Current Biology* 18(6), 458-463.
- Sikorski, T.W. & Buratowski, S. (2009). The basal initiation machinery: beyond the general transcription factors. *Current Opinion in Cell Biology* 21(3), 344-351.
- Silva, J.C., Gorenstein, M.V., Li, G.Z., Vissers, J.P.C. & Geromanos, S.J. (2006). Absolute quantification of proteins by LCMSE – A virtue of parallel MS acquisition. *Molecular & Cellular Proteomics* 5(1), 144-156.
- Somerville, C. (2006). Cellulose synthesis in higher plants. *Annual Review of Cell and Developmental Biology* 22, 53-78.
- Srivastava, V., Srivastava, M.K., Chibani, K., Nilsson, R., Rouhier, N., Melzer, M. & Wingsle, G. (2009). Alternative Splicing Studies of the Reactive Oxygen Species Gene Network in *Populus* Reveal Two Isoforms of High-Isoelectric-Point Superoxide Dismutase. *Plant Physiology* 149(4), 1848-1859.
- Stout, J., Romero-Severson, E., Ruegger, M.O. & Chapple, C. (2008). Semidominant mutations in *Reduced Epidermal Fluorescence 4* reduce phenylpropanoid content in *Arabidopsis*. *Genetics* 178(4), 2237-2251.
- Taylor, G. (2002). *Populus: Arabidopsis* for forestry. Do we need a model tree? *Annals of Botany* 90(6), 681-689.
- Tohge, T., Nishiyama, Y., Hirai, M.Y., Yano, M., Nakajima, J., Awazuhara, M., Inoue, E., Takahashi, H., Goodenowe, D.B., Kitayama, M., Noji, M., Yamazaki, M. & Saito, K. (2005). Functional genomics by integrated analysis of metabolome and transcriptome of *Arabidopsis* plants over-expressing an MYB transcription factor. *Plant Journal* 42(2), 218-235.
- Tomáscaron, R., Yan, L., K, J., rcaron, enková & Foret, F. (2007). Autofocusing and ESI-MS analysis of protein digests in a miniaturized multicompartement electrolyzer. *Electrophoresis* 28(13), 2283-2290.
- Trygg, J. (2002). O2-PLS for qualitative and quantitative analysis in multivariate calibration. *Journal of Chemometrics* 16(6), 283-293.
- Trygg, J. & Wold, S. (2003). O2-PLS, a two-block (X-Y) latent variable regression (LVR) method with an integral OSC filter. *Journal of Chemometrics* 17(1), 53-64.
- Tuskan, G.A., DiFazio, S., Jansson, S., Bohlmann, J., Grigoriev, I., Hellsten, U., Putnam, N., Ralph, S., Rombauts, S., Salamov, A., Schein, J., Sterck, L., Aerts, A., Bhalarao, R.R., Bhalarao, R.P., Blaudex, D., Boerjan, W., Brun, A., Brunner, A., Busov, V., Campbell, M., Carlson, J., Chalot, M., Chapman, J., Chen, G.L., Cooper, D., Coutinho, P.M., Couturier, J., Covert, S., Cronk, Q., Cunningham, R., Davis, J., Degroev, S., Dejardin, A., Depamphilis, C., Detter, J., Dirks, B.,

- Dubchak, I., Duplessis, S., Ehling, J., Ellis, B., Gendler, K., Goodstein, D., Gribskov, M., Grimwood, J., Groover, A., Gunter, L., Hamberger, B., Heinze, B., Helariutta, Y., Henrissat, B., Holligan, D., Holt, R., Huang, W., Islam-Faridi, N., Jones, S., Jones-Rhoades, M., Jorgensen, R., Joshi, C., Kangasjarvi, J., Karlsson, J., Kelleher, C., Kirkpatrick, R., Kirst, M., Kohler, A., Kalluri, U., Larimer, F., Leebens-Mack, J., Leple, J.C., Locascio, P., Lou, Y., Lucas, S., Martin, F., Montanini, B., Napoli, C., Nelson, D.R., Nelson, C., Nieminen, K., Nilsson, O., Pereda, V., Peter, G., Philippe, R., Pilate, G., Poliakov, A., Razumovskaya, J., Richardson, P., Rinaldi, C., Ritland, K., Rouze, P., Ryaboy, D., Schmutz, J., Schrader, J., Segerman, B., Shin, H., Siddiqui, A., Sterky, F., Terry, A., Tsai, C.J., Uberbacher, E., Unneberg, P., Vahala, J., Wall, K., Wessler, S., Yang, G., Yin, T., Douglas, C., Marra, M., Sandberg, G., Van de Peer, Y. & Rokhsar, D. (2006). The genome of black cottonwood, *Populus trichocarpa* (Torr. & Gray). *Science* 313(5793), 1596-1604.
- van de Poppel, J., Kettelarj, N., van Bakel, H., Kockelkorn, T., van Leenen, D. & Holstege, F.C.P. (2005). Mediator expression profiling epistasis reveals a signal transduction pathway with antagonistic submodules and highly specific downstream targets. *Molecular Cell* 19(4), 511-522.
- van den Berg, D.L.C., Snoek, T., Mullin, N.P., Yates, A., Bezstarosti, K., Demmers, J., Chambers, I. & Poot, R.A. (2010). An Oct4-Centered Protein Interaction Network in Embryonic Stem Cells. *Cell Stem Cell* 6(4), 369-381.
- Vaucheret, H. (2006). Post-transcriptional small RNA pathways in plants: mechanisms and regulations. *Genes and Development* 20(7), 759-771.
- Walsh, C.T., Garneau-Tsodikova, S. & Jr., G.J.G. (2005). Protein Posttranslational Modifications: The Chemistry of Proteome Diversifications. *Angewandte Chemie International Edition* 44(45), 7342-7372.
- Weake, V.M. & Workman, J.L. (2008). Histone Ubiquitination: Triggering Gene Activity. *Molecular Cell* 29(6), 653-663.
- Whitelegge, J. (2008). *Protein Mass Spectrometry*: Elsevier Science. (Comprehensive Analytical Chemistry; 52). ISBN 044453055X.
- Wierzbicki, A.T., Ream, T.S., Haag, J.R. & Pikaard, C.S. (2009). RNA polymerase v transcription guides ARGONAUTE4 to chromatin. *Nature Genetics* 41(5), 630-634.
- Wightman, R. & Turner, S.R. (2008). The roles of the cytoskeleton during cellulose deposition at the secondary cell wall. *Plant Journal* 54(5), 794-805.
- Wilkins, O., Nahal, H., Foong, J., Provart, N.J. & Campbell, M.M. (2009). Expansion and Diversification of the *Populus* R2R3-MYB Family of Transcription Factors. *Plant Physiology* 149(2), 981-993.
- Witze, E.S., Old, W.M., Resing, K.A. & Ahn, N.G. (2007). Mapping protein post-translational modifications with mass spectrometry. *Nature Methods* 4(10), 798-806.
- Wold, S., Esbensen, K. & Geladi, P. (1987). Principal Component Analysis. *Chemometrics and Intelligent Laboratory Systems* 2(1-3), 37-52.
- Zhong, R.Q., Demura, T. & Ye, Z.H. (2006). SND1, a NAC domain transcription factor, is a key regulator of secondary wall synthesis in fibers of *Arabidopsis*. *Plant Cell* 18(11), 3158-3170.

- Zhong, R.Q., Lee, C.H. & Ye, Z.H. (2009). Functional Characterization of Poplar Wood-Associated NAC Domain Transcription Factors. *Plant Physiology* 152(2), 1044-1055.
- Zhong, R.Q., Lee, C.H., Zhou, J.L., McCarthy, R.L. & Ye, Z.H. (2008). A Battery of Transcription Factors Involved in the Regulation of Secondary Cell Wall Biosynthesis in *Arabidopsis*. *Plant Cell* 20(10), 2763-2782.
- Zhong, R.Q. & Ye, Z.H. (2007). Regulation of cell wall biosynthesis. *Current Opinion in Plant Biology* 10(6), 564-572.
- Zhou, W., Zhu, P., Wang, J., Pascual, G., Ohgi, K.A., Lozach, J., Glass, C.K. & Rosenfeld, M.G. (2008). Histone H2A Monoubiquitination Represses Transcription by Inhibiting RNA Polymerase II Transcriptional Elongation. *Molecular Cell* 29(1), 69-80.

Acknowledgements

I want to thank my supervisor **Gunnar Wingsle** for drafting me to the proteomics research team. The most striking feature in our group has been the open discussions and freedom to work independently. This approach has given me a good education and by collaboration with other groups we have covered a wide range of research. We have analyzed and studied everything from ants to human but foremost plants. My co-supervisor **Laszlo Bako** has given good advices and I have appreciated our discussions.

My closest colleague from the beginning of my PhD has been **Vaibhav Srivastava**. We started at the same time and begun the work immediately by harvesting a greenhouse full with poplar trees. He introduced me to the Indian culture and I learned a few spicy dishes. The trip to India was arranged in detail and I had a relaxing and amazing experience, thank you so much. His teacher **Manoj S. Srivastava** was a good host at the Bundelkhand University and his work at UPSC was also highly appreciated.

I have experienced a good atmosphere at UPSC; this is largely due to a happy group of coworkers. The photographer, microscopist and all other things, **Kjell** deserves great thanks as a booster of our nice working environment. There are several boosters among the staff that have made the work much more than science. I am thanking **Inga-Britt, Maria, Gun-Britt, Ingela, Inga-Lis, Johanna, Gertrud, Krister, Britt-Marie, Margareta, Roger, Stefan** and **Kjell E**. I value the help and patients **Jan-Erik N** demonstrated making member lists for the PhD council and my own course registration. The final piece of the puzzle providing a good atmosphere is of course **Siv** at Umeå University.

There are several people that I have meet in creative discussion regarding different project. It has become a long list and I can remember **Johan Skog** fishing for

adenovirus receptors, **Stefan Burén** looking deep at carbonic anhydrase and glycosylation, **Mattia P** interested in dinoflagellates grown with different ratios of C and N, **Sara V P** modifying protocols to analyze sorted cells, **Edouard P** interested in secreted proteins during Zinnia TE formation, **Anil** struggling with nuclear extractions, **Linnea M** determining the degradation of the streptococcal IgG cleaving enzyme IdeS, **Pakala S** making isolations of ionically bound xylem cell wall proteins, **Junko** looking for putative xylanases and her supervisor **Ewa** interested in pectin methylesterases, **Zaki** dealing with yeast mediator purifications from glycerol gradients, **Agnieszka** interested in UGPase, **Christine** dealing with the GT43 family, **Emma** having the dear GUTs as favorites, **Manoj K** looking for the invincible MAP20, **Sabine** interested in UGPase and proteins that interact with certain DNA-elements, **Jakob** planning phosphorprotein experiments, **Carl W** and **Lina M** identifying cerebrospinal fluid proteins from SELDI plates, **Maribel** trying to find potential substrates for her favorite protease (metacaspase), **Mikael S** fishing for novel microtubule proteins, **Mikael N** interested in ubiquitinated proteins, **Miriam** making mediator chromatin isolations, **Kamel** expressing MYB proteins in quantities, **Sandra** curious about amino acid compositions in proteins, **David** looking at the fasciclin-like arabinogalactan-protein family, Finally the happy Ukrainian **Sergey** taking on a deep exploration of cellulose synthesis.

I feel gratitude to all fellow PhD students (and a few brave postdocs) that have organized and participated in the traditional spex (amateur comedy) that is arranged at the dissertation party. Many PhD students have devoted a great deal of time to make a memorable finish to our studies. It is impossible for me to forget some grand performances like **Henrik B** acting as Ove and Thomas, **Sara** acting as Markus, **Mattias** acting as Kjell, **Mathieu** in big and small costumes, **Ellinor** acting as Junko. Special tanks to **Maria I** that was keeping the traditions going and arranged many other activities at the beginning of my PhD studies.

Thank you **Marlene** for explaining the earlier work performed in our group and giving advice regarding my PhD education. I remember **Johanna** working in our group in the beginning. You were a colorful spot among us guys, well organized and made an excellent work in introducing me to the lab.

Thank to all dedicated students engaged in the SHS PhD council. It has been a pleasure to participate in the activities and I have learned a lot as a representative in the council and faculty board.

Thank you **Stefan Bäckström** and **Nils** for sincere enthusiasm in our work together, it gave me a good start in my PhD studies. I need to thank **Stefan Björklund** for introducing me to the fascinating mediator and reviewing my thesis.

I want to thank **Katja** for our two success collaborations, first the well coordinated kubb game in Lycksele and secondly the publication. Your supervisor **Christer** deserves my appreciation as a master of membrane proteins. We have had many long discussions on the phone and the correspondence by mail is by far the longest. I got many good ideas and the detailed work was in the end of high quality.

A lot of my work would not be possible without a staff of computer guys. I want to thank **Bo Segerman** for help converting databases, **Andreas Sjödín** spending time making scripts to calculate peptide properties and co-expression networks with interesting results not yet used for any publication, my **brother Andreas** making a program to merge files, **Max** and **Andreas G** in our struggle to handle the large datasets. Bearing a heavy load is **Joakim**, the only one remaining at UPSC maintaining databases and at the same time taking on new tasks.

Thank you **Johan T** for listening to my concerns and taking action to improve the propagation of poplar in the greenhouses. You shown a fast pace in work also reflected in the rate your students take their degrees. I want to thank **Markus** for advices and knowledge regarding submissions to the journals.

I want to thank **Irene** for the nice Christmas tree logging event and inviting our group for dinner at the big ranch. I need to thank my officemates **Vaibhav**, **Delphine**, **Rozen**, **Jon** and **Sara** for enduring my disrupting discussion with students, friends and collaborators and for giving a peaceful working environment.

I want to thank **Gunny**, **Hans** and **Eva** for your interest in my research and your wonderful dinners. Thank to my **cousins**, **aunts** and **uncles** for support and the enjoyable Nilsson BBQ party.

I have really enjoyed my time as a PhD student. It has been interesting to work together with all nationalities and most importantly it made me meet my love. During the coldest winter I found **Rasika** that have totally changed my life. You gave me hope for the future and made me set the date for my defense. Now I am devoted to start a new life and we can do great science together. My achievements would not be possible without my family and relaxation. The best moments during this time have been the work together with my **brother Thomas** on his cottage, the SMS from my **brother Andreas** and **Ylva** that I have become an uncle to a

charming girl, **Hilda**. Thanks to the kind **Grönlund family** for all nice moments.
I want to thank my **mother Inger** and **father Tore**, responsible for my upbringing, for your everlasting support.

As a final point, I have started to like Umeå, but not more than Piteå, Yet!